

Refleksjonsnotat

TDT4111 ANVENDT PROGRAMMERING -- PROSJEKTGRUPPE 50
KATRINE STRANDE RØV, SILJE SALTNES OLSEN & THEA DIGERNES NOSSEN

Læringsutbytte

Prosjektet ga oss innsikt i hele prosessen fra innhenting av miljødata til bearbeiding, visualisering, analyse og enkel prediktiv modellering. Vi lærte å hente værdata via Frost API og bearbeide dem med Pandas og NumPy. Vi håndterte manglende verdier, rensset og strukturerte data, og analyserte sammenhenger mellom temperatur, nedbør og luftfuktighet. Regresjonsanalyse ga erfaring med modelltrening, og visualisering med Matplotlib, Seaborn og Plotly hjalp oss å tolke resultatene. Dette styrket vår forståelse av datavitenskapelige metoder i praksis, og hvordan slike metoder kan anvendes på reelle og samfunnsrelevante problemstillinger.

Nye ferdigheter

Vi tilegnet oss ferdigheter i datainnhenting, strukturering og analyse med Pandas og NumPy. Vi lærte å bruke visualiseringsverktøy og fikk erfaring med prediktiv modellering og evaluering med R^2 -score og scatterplots. Vi testet også nye verktøy som Plotly og missingno, som ga oss bedre innsikt i datakvalitet og visualisering.

En viktig læring var bruk av Git og GitHub. Vi hadde lite erfaring med versjonskontroll og støtte på utfordringer med commits, push og pull. Etter tilbakemelding på del 1 gjorde vi strukturelle endringer og opprettet et nytt repo. Dette lærte oss viktigheten av hyppige commits, gode meldinger og jevnlig oppdatering av lokal kodebase. Det var en bratt læringskurve, men ga verdifull innsikt i samarbeid i kodeprosjekter og hvordan man kan jobbe effektivt i team.

Utfordringer

Vi støtte på flere utfordringer. I starten fant ikke API fra MET, da vi lette på feil nettside, men fikk hjelp av en annen gruppe. Det tok tid å finne værstasjoner med tilstrekkelige data, og vi fikk ikke lastet inn vinddata. Nedbørdata viste noen uventede verdier, noe som krevde ekstra kvalitetssikring.

En tidlig feil var hardkoding av API-nøkkel. Etter tilbakemelding lærte vi å bruke .env-filer og dotenv-biblioteket, og byttet nøkkelen.

Vi brukte tid på å forstå strukturen i responsdataene fra Frost. Tekniske problemer oppsto også mellom Windows og Mac, spesielt med filstier og encoding. Manglende erfaring med moduloppbygging, testing og dokumentasjon førte til at vi måtte lære mye underveis. Etter del 1 opprettet vi et nytt Git-repo etter råd fra studentassistent, da den opprinnelige

strukturen var uoversiktlig. Koden ble kopiert over, slik at vi kunne fortsette uten å miste fremdrift. Dette ga oss en ny start og bedre kontroll over prosjektets videre utvikling.

Samarbeid

Vi jobbet samlet gjennom store deler av prosjektet, noe som fungerte godt for problemløsning og fremdrift. Oppgaver ble fordelt løpende, men vi samarbeidet ofte på tvers.

Kommunikasjonen var god, og vi lærte mye av å diskutere problemer sammen. I ettertid ser vi at tydeligere rollefordeling og mer strukturert arbeidsflyt kunne vært nyttig, særlig for testing og dokumentasjon. Samtidig var det en styrke at vi løste vanskelige oppgaver sammen og holdt høy aktivitet gjennom hele prosjektet. Det kollektive engasjementet bidro til at vi opprettholdt motivasjonen og lærte mye av hverandre.

Vurdering av resultatene

Vi er fornøyde med å ha hentet og analysert et stort datagrunnlag for tre byer over mer enn to år. Visualiseringene viser tydelige trender for temperatur, luftfuktighet, nedbør og vind.

Bruken av missingno ga innsikt i datakvaliteten før og etter rensing. Regresjonsmodellene forklarer variasjon i luftfuktighet basert på temperatur og nedbør. Resultatene ble visualisert med scatterplots og evaluert med R^2 -score og MSE, som viser at modellene fungerer godt.

Koden er strukturert i moduler, med dokumentasjon og enhetstester i henhold til kravene. Vi opplever at sluttproduktet er både funksjonelt og faglig solid.

Forbedringsforslag

En utfordring var uklar oppgavetekst og varierende svar fra studentassistenter, noe som skapte usikkerhet. En tydeligere kravspesifikasjon og enhetlig veiledning ville gjort prosessen mer forutsigbar og læringsutbyttet bedre. Samtidig satte vi pris på muligheten til å bruke kreativitet i prosjektet.

Vi mener også at tidsfordelingen mellom del 1 og del 2 var skjev. Del 1 hadde vi god tid på, mens del 2 ble presset inn på kortere tid. Fristen for siste veiledning kom tidligere enn forventet. Vi foreslår at del 1 leveres tidligere, slik at det blir bedre tid til å ferdigstille del 2. Tilbakemeldingen på del 1 opplevdes også som overfladisk og lite nyttig, og vi savnet mer konkret veiledning på forbedringspunkter.

Videre utvikling

Prosjektet kan videreutvikles ved å hente sanntidsdata, bruke mer avanserte modeller som beslutningstrær eller nevrale nettverk, og implementere geografisk visualisering. Automatisert

datainnhenting og et mer interaktivt brukergrensesnitt ville også vært nyttig. Flere enhetstester og bruk av logging i stedet for print-utskrifter ville økt robustheten. Det er også potensial for å gjøre prosjektet mer skalerbart og tilgjengelig for andre brukere.

Oppsummering og fremtidig nytte

Prosjektet har gitt oss praktisk erfaring med hele dataanalyseprosessen – fra innhenting til modellering. Vi har lært å strukturere kode, samarbeide effektivt og dokumentere prosess og resultat. Erfaringen med Git, testing og API-bruk har høy overføringsverdi til videre studier og arbeidsliv. Vi har fått en dypere forståelse for hvordan datavitenskap kan brukes til å analysere reelle problemstillinger og gi innsikt som er relevant for både fag og samfunn. Dette prosjektet har vært en verdifull læringsarena og et godt grunnlag for videre utvikling innen datavitenskap.