

Vedlegg 3 – Dokumentasjon og innlevering

Dokumentasjon

Deltakere i denne datakonkurransen skal levere dokumentasjon for løsningen, kildekoden i tekstlig format (Word, PDF, TXT)

a. Metodevalg og tilnærming

Denne seksjonen skal forklare hvilken metodikk som ble brukt og begrunne hvorfor metoden er egnet for å løse problemstillingen i datakonkurransen.

b. Systemstruktur og arkitektur

Beskriv hvordan løsningen er bygget opp teknisk, inkludert komponenter og eventuelle API-er. Her anbefales en visualisering av systemarkitektur.

c. Modeller og algoritmer

Dokumenter hvilke modeller som er brukt (f.eks. maskinlæring, statistiske modeller). Beskriv parametervalg og evalueringsmetoder.

d. Kildekode

Gi oversikt over kodebasen, med tilhørende kommentarer og forklaringer. Beskriv hvordan løsningen kan installeres, kjøres og evt. videreutvikles eller skaleres.

Innlevering og evaluering

Format

Prediksjonene skal leveres som tekstfil i csv-format, UTF-8. Filen skal inneholde følgende kolonner, separert med komma:

airport_group - tekstfelt med en bokstav som representerer flyplassgruppe

date - tekstfelt med dato prediksjonen gjelder for med formatet YYYY-MM-DD

hour - integer med timen som prediksjonen gjelder for - f.eks. verdien 9 gjelder for tidsrommet 09:00-09:59

pred - sannynlighet for samtidighet med opptil 3 desimaler - obs, bruk punktum som desimaltegn.

Et eksempel på hvordan det vil skal ut:

```
airport_group,date,hour,pred
```

```
A,2025-10-01,9,0.812
```

```
B,2025-10-01,9,0.113
```

```
A,2025-10-01,10,0.420
```

```
B,2025-10-01,10,0.982
```

Dersom dere ønsker kan filen ***preds_mal.csv*** brukes som mal. Her har vi generert alle kombinasjoner av flyplassgrupper og tider som vi forventer. Verdiene i kolonnen pred skal bytte ut. Den er forhåndsutfyllt med verdiene 0.0. Gi filen et navn som hjelper oss å identifisere innsender, f.eks. etternavn.csv

Evaluering

Etter at perioden vi ønsker å predikere for har passert kommer vi til å generere en tabell med faktiske verdier basert på observerte samtidighetene i perioden. Denne kommer til å ha dette formatet:

airport_group,date,hour,actual

A,2025-10-01,9,1

B,2025-10-01,9,0

A,2025-10-01,10,1

B,2025-10-01,10,0

Innleveringene slås sammen med fasiten slik:

airport_group,date,hour,pred,actual

A,2025-10-01,9,0.812,1

B,2025-10-01,9,0.113,0

A,2025-10-01,10,0.420,1

B,2025-10-01,10,0.982,0

Det er avgjørende at nøklene i bidragene (kombinasjonen av airport_group, date og hour) samsvarer med nøklene i fasiten. Hvis nøklene ikke stemmer overens, vil bidraget ikke kunne evalueres.

Evalueringsmetode

Vi beregner AUC-ROC for å finne de beste prediksjonene.