

# Vedlegg 1: Beskrivelse av datasett

I dette vedlegget beskrives datasettene som deles for å ta i bruk i konkurransen. Datakonkurransen bruker følgende datasett:

Oppgaven går ut på å bygge en maskinlæringsmodell som kan predikere sannsynlighet for samtidighet pr flyplassgrupper pr time i perioden 01.10.2025 til 31.10.2025.

## Datagrunnlag

### *Historisk ruteprogram*

Datasett med flybevegelser, planlagte og faktiske avgangs- og landingstider, flyplasser, flyselskap, flyvningstype, mm. Datasettet utleveres som CSV.

**Fil:** historical\_flights.csv

*flight\_id* – flyvningsnummer, første to tegn identifiserer flyselskapet

*dep\_airport* – avgangsflyplass

*dep\_airport\_group* – gruppe som avgangsflyplassen tilhører

*arr\_airport* – ankomstflyplass

*arr\_airport\_group* – gruppe som ankomstflyplassen tilhører

*service\_type* – flyvingsart, f.eks. rute, charter, frakt, osv

*cancelled* – 0/1 flagg som indikerer om planlagt flyvning ble kansellert

*std* – planlagt avgangstid (*scheduled time of departure*)

*sta* – planlagt ankomstid (*scheduled time of arrival*)

*atd* – faktisk avgangstid (*actual time of departure*)

*ata* – faktisk ankomstid (*actual time of arrival*)

### *Treningsdata*

Vi har brukt datasettet over (historical\_flights.csv) til å lage et enkelt datasett for trening. I vårt treningssett har vi laget et par enkle features til inspirasjon. Vi oppfordrer deltakere til å lage egne features, både basert på ruteprogrammet og på andre åpne datakilder, for eksempel værdato, helligdager, og annet. Vi tror at det er mye å hente i å være kreativ her.

**Fil:** training\_data.csv

*airport\_group* – se tabellen *airportsgroups.csv* for å se hvilke flyplasser hører til hvilken gruppe

*date* - dato

*hour* - time, f.eks. gjelder verdien 9 for tidsrommet 09:00-09:59

*target* – 0/1-flagg som indikerer om det har skjedd samtidighet basert på faktiske flytider

*feat\_season* – årstid

*feat\_sched\_concurrence* – 0/1-flagg for antatt samtidighet basert på planlagte flytider

### *Prediksjonsdata*

Foreløpig ruteprogram for oktober 2025 legges ut sammen med oppgaven.

**Obs!: Oppdatert ruteprogram for oktober vil legges ut 26.september, kl. 09.00.**

**Fil:** schedule\_oct2025.csv

*flight\_id* – flyvningsnummer, første to tegn identifiserer flyselskapet

*dep\_airport* – avgangsflyplass

*dep\_aiport\_group* – gruppe som avgangsflyplassen tilhører

*arr\_airport* – ankomstflyplass

*arr\_airport\_group* – gruppe som ankomstflyplassen tilhører

*service\_type* – flyvingsart, f.eks. rute, charter, frakt, osv

*std* – planlagt avgangstid (*scheduled time of departure*)

*sta* – planlagt ankomstid (*scheduled time of arrival*)

Dette datasettet kan brukes til å lage data som dere skal gjøre prediksjon på. Vi har laget et eksempel på dette her:

**Fil:** inference\_data\_oct2025.csv

*airport\_group* – se tabellen *airportsgroups.csv* for å se hvilke flyplasser hører til hvilken gruppe

*date* - dato

*hour* - time, f.eks. gjelder verdien 9 for tidsrommet 09:00-09:59

*target* – 0/1 flagg som indikerer om det har vært samtidighet i det aktuelle tidsrommet

*feat\_season* – årstid

*feat\_sched\_concurrence* – 0/1-flagg for antatt samtidighet basert på planlagte flytider