

# EO-Ø2-06 2024 Vedlegg B-6 Statistisk modell

**Dokumentet er signert digitalt av følgende undertegnere:**

- MARGARETH NORDBY-KRINGLI, signert 20.12.2024 med ID-Porten: BankID
- ERIK LUND, signert 20.12.2024 med ID-Porten: BankID



## **Det signerte dokumentet inneholder**

- En forside med informasjon om signaturene
- Alle originaldokumenter med signaturer på hver side
- Digitale signaturer



## **Dokumentet er forseglet av Posten Norge**

Signeringen er gjort med digital signering levert av Posten Norge AS. Posten garanterer for autentisiteten og forseglingen av dette dokumentet.



## **Slik ser du at signaturene er gyldig**

Hvis du åpner dette dokumentet i Adobe Reader, skal det stå øverst at dokumentet er sertifisert av Posten Norge AS. Dette garanterer at innholdet i dokumentet ikke er endret etter signering.

## Trafikkavtale Østlandet 2

- Regionekspresstog (RE-tog) Skien-Eidsvoll
- Regionekspresstog (RE-tog) Drammen-Lillehammer
- Regiontog (R-tog) Kongsberg - Eidsvoll
- Regiontog (R-tog) Drammen - Dal
- Regiontog (R-tog) Asker - Kongsvinger
- Regiontog (R-tog) Notodden - Porsgrunn (Bratsbergbanen)
- Regiontog (R-tog) Drammen - Oslo lufthavn
- Regiontog (R-tog) Stabekk/Oslo S - Oslo lufthavn

E0-Ø2-06 2024 Vedlegg B-6 Statistisk modell



Dokumentet er signert digitalt av:

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

Forseglet av



Posten Norge

**Endringslogg**

| Versjon | Hva                                         | Dato       |
|---------|---------------------------------------------|------------|
| 1.0     | EO-Ø2-06 2024 Vedlegg B-6 Statistisk modell | 20.12.2024 |
|         |                                             |            |
|         |                                             |            |
|         |                                             |            |

Dokumentet er signert digitalt av:

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

Forseglet av



Posten Norge

# ENDRINGSORDRE

|                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                       |                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| Avtalenr.:<br>Ø2                                                                                                                                                                                                                                                                                                                                                                                                                | Endringsordre nr.:<br>06                                                                                                                                              | Antall vedlegg:<br>1 |
| Ref. Anmodning om endringsordre (AOE) nr.:                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                       |                      |
| Beskrivelse av endring (inklusive referanse til alle aktuelle vedlegg):<br><br>Trafikkavtalen har fått et oppdatert Vedlegg B-6 Statistisk modell. Oppdateringen er for at det skal være mer i samsvar med hvordan modellen fungerer, dette gjelder beregninger med avvikstransport og kalibrering, videre er det gjort en liten jobb med å tydeliggjøre de andre punktene.<br><br>Se vedlegg 1: Vedlegg B-6 Statistisk modell. |                                                                                                                                                                       |                      |
| Kostnader av endring:<br><br>N/A                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                       |                      |
| Betalingsform for endring:<br><br>N/A                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                       |                      |
| Endring av tidsplan/milepæler (dersom endringen ikke har konsekvenser for tidsplan, skriv "ingen"):<br><br>N/A                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                       |                      |
| Oppstarttidspunkt for endring:<br><br>Umiddelbart.                                                                                                                                                                                                                                                                                                                                                                              | Ferdigstillelse av endring:<br><br>Endringen følger trafikkavtalens varighet.                                                                                         |                      |
| Leverandøren instrueres herved om omgående å igangsette Endring av Arbeidet som beskrevet ovenfor. I den utstrekning denne Endringsordre ikke bestemmer annet, gjelder bestemmelsene i Trafikkavtalen og tidligere Endringsordre.                                                                                                                                                                                               |                                                                                                                                                                       |                      |
| Ordre utstedt dato:<br><br>Navn: Erik Lund<br><br>Tittel: Direktør avdeling for kjøp og avtalestyring<br><br>Signatur:<br><br><div style="text-align: right;">Jernbanedirektoratet.</div>                                                                                                                                                                                                                                       | Ordre akseptert dato:<br><br>Navn: Margareth Nordby-Kringli<br><br>Tittel: Regiondirektør<br><br>Signatur:<br><br><div style="text-align: right;">Leverandøren.</div> |                      |



Dokumentet er signert digitalt av:

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

Forseglet av



Posten Norge

## Trafikkavtale Østlandet 2

- Regionekspresstog (RE-tog) Skien - Eidsvoll
- Regionekspresstog (RE-tog) Drammen - Lillehammer
- Regiontog (R-tog) Kongsberg - Eidsvoll
- Regiontog (R-tog) Drammen - Dal
- Regiontog (R-tog) Asker - Kongsvinger
- Regiontog (R-tog) Notodden - Porsgrunn (Bratsbergbanen)
- Regiontog (R-tog) Drammen - Oslo lufthavn
- Regiontog (R-tog) Stabekk/Oslo S - Oslo lufthavn

### Vedlegg B-6 Statistisk modell 2.0



Dokumentet er signert digitalt av:

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

Forseglet av



Posten Norge

**Endringsloggen**

| Versjon | Hva                                                                                                            | Dato       |
|---------|----------------------------------------------------------------------------------------------------------------|------------|
| 1.0     | Vedlegg B-6.                                                                                                   | 29.06.2023 |
| 2.0     | Oppdatert tekst til punktene 1, 1.1, 2.2, 3.1, 3.2, 3.2.2, 3.3.1, 3.4.1, 3.4.2, 3.4.3, 3.4.4, 3.4.5, 3.5 og 5. | 05.11.2024 |

Dokumentet er signert digitalt av:

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

Forseglet av



Posten Norge

# Innhold

|                                                              |           |
|--------------------------------------------------------------|-----------|
| <b>1 Innledning.....</b>                                     | <b>4</b>  |
| 1.1 Bakgrunn.....                                            | 4         |
| <b>2 Formål og grunnprinsipper .....</b>                     | <b>4</b>  |
| 2.1 Formålet med modellen.....                               | 4         |
| 2.2 Grunnprinsipper .....                                    | 4         |
| <b>3 Oversikt over prosessen .....</b>                       | <b>5</b>  |
| 3.1 Innledning .....                                         | 5         |
| 3.2 Datagrunnlaget (inputdata) .....                         | 5         |
| 3.2.1 Primærdata .....                                       | 5         |
| 3.2.2 Sekundærdata .....                                     | 6         |
| 3.3 Korreksjoner, vasking og beriking av Datagrunnlaget..... | 6         |
| 3.3.1 Vasking av Telledata .....                             | 6         |
| 3.3.2 Produksjonsdata .....                                  | 8         |
| 3.4 Behandling av data i statistisk modell.....              | 8         |
| 3.4.1 Prosessering av data .....                             | 8         |
| 3.4.2 Retrening .....                                        | 8         |
| 3.4.3 Kalibrering .....                                      | 9         |
| 3.4.4 Konduktørjustering .....                               | 9         |
| 3.4.5 Beregning av passasjertall i avvikssituasjon .....     | 9         |
| 3.4.6 Utdata fra statistisk modell .....                     | 9         |
| <b>4 Tilgang til data .....</b>                              | <b>10</b> |
| <b>5 Endringer og justeringer i modellen.....</b>            | <b>10</b> |



# 1 Innledning

## 1.1 Bakgrunn

Trafikkinntektene på linjer som driftes av Leverandøren skal fordeles gjennom en inntektsfordelingsmodell (jf. «Vedlegg B-7») som baserer seg på totalt antall Passasjerer fra de ulike togproduktene.

Oppdragsgiver har utviklet en statistisk modell som skal supplere tellinger med estimater for å sikre et best mulig bilde av totalt antall Passasjerer. Modellen skal estimere Passasjertall der Telldata mangler.

Den statistiske modellen er utviklet basert på historiske APC-tellinger, Produksjonsdata og Sekundærdata. Utviklingen av modellen og datagrunnlaget som er benyttet i utviklingen er nærmere beskrevet i Modelldokumentasjonen.

Den statistiske modellen driftes og vedlikeholdes av Oppdragsgiver.

# 2 Formål og grunnprinsipper

## 2.1 Formålet med modellen

Den overordnede målsetningen med den statistiske modellen er å sikre et så presist som mulig og dermed rettferdig, transparent og troverdig passasjergrunnlag.

## 2.2 Grunnprinsipper

For å oppnå det overordnede formålet (i punkt 2.1) gjelder følgende grunnprinsipper for statistisk modell:

1. Statistisk modell utvikles for å gi mest mulig presise resultater til bruk i inntektsfordelingen
  - a. Med «resultater» menes modellens utregning av antall Passasjerer på en gitt linje
  - b. Med «mest mulig presis» menes det at modellens snittfeil er så lav som mulig i en måleperiode (per måned og per linje). Snittfeilen beregnes med Root Mean Square Error (RMSE), Mean Biased Error (MBE) og Mean Average Error (MAE) som alle er statistiske metoder for å måle presisjonen i statistiske estimater.
2. Presisjonsnivået i estimatene kontrolleres månedlig mot faktiske tellinger fra APC-utstyr. Togoperatør vil få tilgang til resultatene.
3. Justeringer i statistisk modell implementeres kun dersom endringen ventes å gi lik eller forbedret presisjon
4. Planlagte justeringer i statistisk modell kommuniseres til Leverandøren dersom de ventes å ha en vesentlig effekt på resultatene



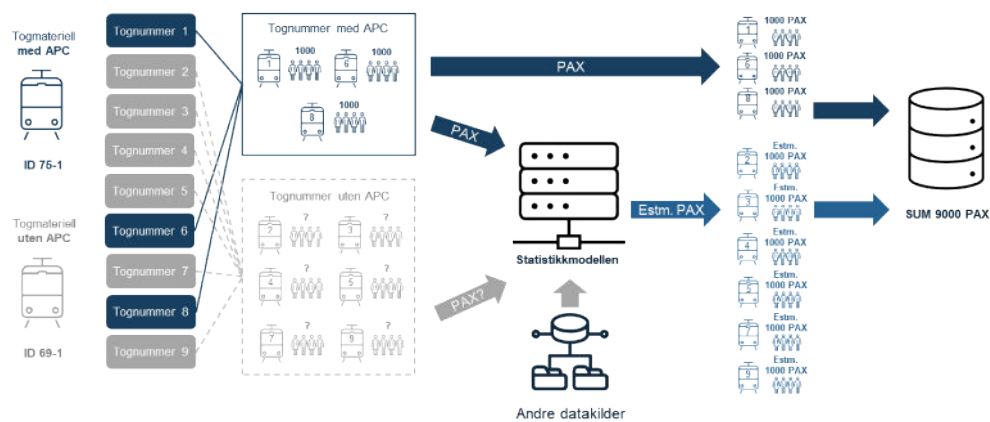
## 3 Oversikt over prosessen

### 3.1 Innledning

Proessen som Oppdragsgiver benytter for å estimere antall Passasjerer er i korte trekk følgende:

1. Oppdragsgiver innhenter datagrunnlaget som togoperatør tilgjengeliggjør (punkt 3.2)
2. Data vaskes og berikes (punkt 3.3)
3. Data behandles i statistisk modell ved at kunstig intelligens foreslår et regelsett for å estimere reiseaktivitet. Regelsettet implementeres i statistisk modell etter kvalitetssikring av Oppdragsgiver. For utfyllende beskrivelse henvises det til Modelldokumentasjonen. Prosessen er illustrert i figuren under:

Figur 1: Illustrasjon av den statistiske modellen



### 3.2 Datagrunnlaget (inputdata)

Datagrunnlaget (Inputdata) som benyttes i den statistiske modellen består av primærdata (Telledata og Produksjonsdata) og Sekundærdata.

Leverandørs leveranse av inputdata til Oppdragsgiver følger av Trafikkavtalenes bestemmelser avsnitt 6.7 (levering og bruk av telledata og annen rådata) og avsnitt 6.8 (Rapporterings- og informasjonsplikt) samt Trafikkavtalens vedlegg A - punkt 6.5 (levering av telledata og annen rådata)

#### 3.2.1 Primærdata

Modellen benytter Telledata og Produksjonsdata som de sentrale inputkilder for å beskrive togtrafikken.

##### Telledata (fra APC-utstyr)

Telledata inneholder opplysninger om antallet på- og avstigninger på stasjonen, per dør, per Togsett, per stasjon, per tognummer. Basert på antall Påstigende og Avstigende, kan vi også utregne PAX på et gitt tog. APC-dataene utgjør dermed fundamentet for input til modellen.

### *Produksjonsdata*

Produksjonsdata gir informasjon om alle tog, også togene som ikke har APC-utstyr.

Produksjonsdata benyttes som grunnlag for hvor og når det kjører et tog, og hvilke tog som kjører. Dette er avgjørende for å estimere antall Passasjerer.

Produksjonsdataene gir også opplysninger om produksjonen, samt flere forskjellige variabler relatert til produksjonen, f.eks. forsinkelse, innstillinger, lokasjon, avgangstid, kapasitet osv.

### **3.2.2 Sekundærdata**

Sekundærdata som benyttes kommer fra ulike tredjeparter. Datakildene som er benyttet i utviklingen av modellen er beskrevet i Modelldokumentasjonen. Eksempler på sekundærdata er værdata, trafikkdata, kalenderinformasjon og områdedata/geografi.

#### Andre sekundærdata/datakilder

Ved behov (eksempelvis ved bortfall av en datakilde) vil det gjøres vurderinger om det er nødvendig å benytte andre Sekundærdata enn de opprinnelig benyttet for å utvikle modellen, og hvilke datakilder som skal benyttes.

Sekundærdata kan hentes fra andre og/eller nye kilder dersom Oppdragsgiver finner det hensiktsmessig, f.eks. at eksisterende datakilder ikke er tilgjengelig.

### **3.3 Korreksjoner, vasking og beriking av Datagrunnlaget**

Før Telledata benyttes for beregning av antall Passasjerer, er det derfor nødvendig å vaske dataene for å forbedre datakvaliteten slik at antall Passasjerer blir beregnet mest mulig presist.

For å ivareta kvaliteten, gjennomføres følgende steg for å berike og vaske datasettet før det benyttes i statistisk modell:

1. Telledata mottas urørt fra APC-utstyr.
2. Telledata vaskes automatisk basert på et gitt regelsett.
3. Telledata og Produksjonsdata komplementeres med Sekundærdata.
4. Kontrollmekanisme validerer tallene (for eksempel ekstremverdier).
5. Det vaskede og berikede datasettet benyttes i modellen for estimering.

For utfyllende beskrivelsen av punktene henvises det til modelldokumentasjon.

#### **3.3.1 Vasking av Telledata**

##### *Utjevning av antall Passasjerer*

En grundig gjennomgang av Telledata synliggjør utfordringer med datasettet som må løses. Dette skyldes to forhold:

- 1) Målefeil: All sensordata vil alltid inneholde en viss grad av feil – dvs. at systemene noen ganger teller en påstigende passasjer i stedet for to. Dersom systemet teller antall avstigende på korrekt måte betyr dette at vi vil ha negativ PAX på toget siste stopp.



- 2) Kunstig ulogisk oppførsel fra de reisende: Eksempelvis forventes det ikke avstiginger på første stopp, men dette forekomme eksempelvis hvis man har hjulpet noen inn på toget. På- og avstiginger fra ombordpersonalet er en annen kilde til slike feil.

(Målefeil og) ulogiskheter i av- og påstigningsmønster forventes å være tilfeldig på tvers av tog. Målefeilen vil allikevel sannsynligvis være korrelert med ytre faktorer som f.eks. passasjertall. Dermed kan målefeilen være større i rushtid enn utenfor rushtid og på større stasjoner sammenlignet med mindre stasjoner.

Begge disse typene feil håndteres ved at de internaliseres i dataene. Dette vil si at dataene endres ved at de må oppfylle et sett med regler. Dette fører til en utjevning av antall passasjerer over togets reise.

En sentral operasjon er nullstilling av togene slik at antall Passasjerer på toget ikke er negativ ved slutten av reisen eller under reisen eller at det ikke er Passasjerer som aldri går av.

Denne nullstillingen gjøres i følgende rekkefølge; først gjøres det opp differanse ved endt tur, for å så fordele differansen ut over hele turen proporsjonalt med antallet av telte Passasjerer ved stoppet.

Dette sikrer at hver tur slutter med null Passasjerer. Har vi et negativt antall Passasjerer ombord underveis, utlignes dette ved å tilføye Påstigende og Avstigende Passasjerer. For utfyllende beskrivelse av prosessen henvises det til Modelldokumentasjonen avsnitt 1.0.

Hvis det ikke er mulig å utjevne hensiktsmessig, betrakter vi tellingene som utroverdige og fjerner disse dataene fra utvalget som benyttes i modellen.

Dersom det oppdages endringer i forutsetningene, vil metodikken og reglene for utjevning endres.

#### *Korrigerende og forenklinger av Telledata*

Telledata som inneholder observasjoner som ikke oppfyller de logiske regler som er beskrevet nedenfor fjernes.

1. **For å unngå dobbelttelling kan en observasjon kun inngå én gang i modellen.** Derfor fjernes reiser hvor det er flere APC-tellingene på samme stopp på samme avgang. For å unngå dobbelttelling kan en observasjon kun inngå én gang i modellen.
2. **Avganger som er innstilt helt fjernes fullstendig.** Dette ettersom vi da vet at det ikke eksisterer noen Passasjerer. Reiser hvor enkelte stopp er innstilt beholdes, ettersom vi ellers ville vært nødt til å fjerne store mengder informative data. Det innstilte stoppet fjernes siden vi her vet at ingen går hverken på eller av.
3. **Reiser som ikke er komplette fjernes.** Det vil si at en reise må ha et første og siste stopp, samt alle stoppene imellom (bortsett fra innstilte stopp). For disse ufullstendige reisene er det ikke mulig å avgjøre hvorfor et stopp mangler. Derfor kan vi ikke vite hvor mange av- og påstigninger det er på stoppet. For å sikre kvalitet i modellen, blir disse reisene frasortert under trening av modellen, men blir tatt tilbake når modellen benyttes til prediksjon av Passasjertall i output. Dette for å sikre høyest mulig dekningsgrad.

For avganger med togsett med APC-utstyr hvor det er mottatt APC-tall på deler av strekningen, men ikke for alle stasjoner, vil Jernbanedirektoratet løpende vurdere og mulig implementere tiltak for bruk av tilgjengelig APC. Slike tiltak vil kun implementeres om de vurderes til å heve presisjonen i modellen og forutsettes at det er mulig å ha tilstrekkelig gode kvalitetssikringsmekanismer, som unngår at støy fra APC skaper vesentlige utslag.



4. **Vi samler det gitte togets dører i par.** Vi vet ikke på hvilken side av toget døren er åpen. Ved å samle dørene i par, gjør vi datasettet halvparten så stort.
5. **Vi vet ikke hvilken vei toget kjører.** Dette betyr at vi ikke vet om dør 1 og 2 er de første 2 dørene som kjører inn på stasjonen, eller de siste. For å ikke skape unødige mye støy i modellen, koder vi dørene sammen, slik at vi identifiserer de fremste og siste dørpar som ender, og de midterste dørrnumrene som midtdører.
6. **Passasjerer kan ikke bevege seg mellom to Togsett.** I utregningen av på- og Avstigende per Tog-leg antas det imidlertid at Passasjerene ikke kan bevege seg mellom de forskjellige Togsettene, og analysen foretas derfor på Togsettnivå. For noen avganger forekommer det derfor flere Togsett (vehicles) i samme tog.

### 3.3.2 Produksjonsdata

Før Produksjonsdata benyttes, kan det i enkelte tilfeller være nødvendig å fylle ut enkelte hull i dataene (f.eks. kan dette være manglende tidsangivelser eller kjøretøytype). For deretter å matche dette med andre kilder.

Følgende regler benyttes for behandling av Produksjonsdata som grunnlag for å utfylle enkelte hull i dataene. Disse prosedyrene sikrer at dataene har nødvendig kvalitet, før de blir brukt til estimering.

- I dataene forekommer det at ankomsttiden er etter avgangstiden, og at det er manglende tidsangivelser. Vi fyller hullene ut ved å estimere kjøretidene mellom stasjonene. Hvis det er en sammenheng med de manglende tidsangivelsene og for eksempel forsinkelse, så vil vi kunne underestimere antallet forsinkede tog, som igjen vil kunne påvirke modellen.
- For å finne antall dører og kjøretøy for avganger, må man vite hvilken kjøretøytype og hvor mange kjøretøy som håndterer avgangen. Datakilden for dette er data om togsammensetning som inneholder Dato og Tognummer, samt modell og antall kjøretøy. I en del tilfeller vil man ikke få treff i denne listen pga. hull i datagrunnlaget. På bakgrunn av datakilden for togsammensetning, er det derfor laget en mapping basert på hvilke togsammensetninger som er vanligst per TogNummer (uten dato). Denne brukes så for å avgjøre modell og antall kjøretøy for avganger som mangler faktisk avgang i datakilden for togsammensetning. I de tilfeller hvor det fortsatt ikke finnes togsammensetningsdata på tognummer-nivå, brukes tilsvarende mapping på linjenivå

## 3.4 Behandling av data i statistisk modell

### 3.4.1 Prosessering av data

All Produksjonsdata og Sekundærdata prosesseres for å lage et sett variabler som er enkle for den statistiske modellen å jobbe med. Utformingen av disse variablene kalles "Feature Engineering", og beskrives nærmere i Modelldokumentasjonen.

### 3.4.2 Retrening

En Retrening vil si at regelsettet oppdateres i lys av nytt datagrunnlag samlet inn siden siste Retrening. Det betyr at en Retrening eksempelvis kan bety at modellen oppdateres for å bedre beskrive nye reisemønster som har blitt observert i befolkningen. Dette gjøres for å sikre at modellen fortsatt predikerer på best mulig måte. Hvor ofte modellen Retrenes i fremtiden vil avhenge av (a) endring i presisjonsnivå i estimeringene og (b) konsekvensene av ytre faktorer som f.eks. store endringer i togtilbud, større kommunesammenslåinger og pandemier.

Hvor ofte modellen Retrenes i fremtiden vil avhenge av (a) endring i presisjonsnivå i estimeringene og (b) konsekvensene av ytre faktorer som f.eks. store endringer i togtilbud.



### 3.4.3 Kalibrering

For å sikre at estimatene er så riktige som mulig er det innført en teknikk i modellen som muliggjør kalibrering av estimerte passasjertall mot gyldige APC-tall. APC-tall er den mest tilgjengelige og pålitelige kilden til informasjon, og derfor benyttes dette datagrunnlaget i kalibrering av estimatene.

Kalibreringen har vist seg å føre til betydelig reduksjon i skjevheten til estimatene. Ved å innføre denne metoden har risikoen for at nye modeller fører til et nivåskifte i passasjertall blitt betydelig redusert.

Modellen vil kun kalibrere ved behov.

### 3.4.4 Konduktørjustering

For å hindre at konduktørens på- og avstigninger bidrar til kunstig høye passasjertall vil det gjøres en nedjustering av antall påstigninger og avstigninger i et forsøk på å minimere effekten av konduktørens aktivitet. Det er usikkerhet knyttet til akkurat hvor mange påstigninger og avstigninger som stammer fra konduktøren. Konduktørjusteringen i Statistisk modell baserer seg derfor på en forutsetning om at én konduktør per togsett går på første stasjon, av på siste stasjon og av og på per stopp imellom. Justeringen skal utligne disse, i den grad justeringen ikke fører til negative verdier for påstigninger, avstigninger og antall passasjerer på toget mellom to stopp.

### 3.4.5 Beregning av passasjertall i avvikssituasjon

Statistisk modell beregner passasjertall også ved bruk avvikstransport (buss for tog).

- Buss for tog estimeres som om det var opprinnelig avgang med tog som gikk
- Dersom det går buss for tog i parallell med tog, vil det ikke estimeres passasjertall for bussen

Statistisk modell er tilpasset slik at man kan benytte APC-tall der hvor dette eksisterer for en delavgang med tog. Modellen estimerer også passasjertall ved avvikstransport hvor de reisende blir henvist til andre kollektivaktørers transportmidler – eksempelvis T-banen i Ruter-området. Disse tallene brukes dog kun til inntektsfordeling mellom Østlandet 1 og Østlandet 2 og vil ikke inngå i totalt antall passasjerer på jernbanen. Passasjerkilometer som produseres på kollektivaktørene sine transportmidler ved avvik kommer til fratrekk i oppgjøret i takstsamarbeidsavtalen.

### 3.4.6 Utdata fra statistisk modell

Følgende nøkkeltall er output fra statistisk modell:

1. Antall Passasjerer om bord i toget etter passasjerutveksling.
2. Brutto Påstigende per tog.
3. Netto passasjerutveksling per tog.
4. Brutto Avstigende per tog.
5. Netto passasjerutveksling pr. dør.

Netto passasjerutveksling pr dør beregnes ut fra en fordelingsnøkkel. Fordelingsnøkkelene er basert på historiske Telledata for hvordan Passasjerer fordeler seg mellom dørene i tog. Da fordelingsmønstrene varierer mellom togtyper og stasjoner, beregnes fordelingsnøkkelene for hver enkel kombinasjon av Tog-leg og Stasjon.

Modelldokumentasjonen beskriver hvordan statistisk modell benyttes for å beregne nøkkeltallene.



Modellen benytter to separate maskinlæringsmodeller for å beregne Passasjerer på togene. Den ene modellen estimerer antall Påstigende. Den andre modellen estimerer antall Passasjerer. Antall Påstigende og Passasjerer i toget estimeres på Togsettnivå. For tog som består av flere Togsett, beregnes antall Passasjerer i toget som summen av estimatene i hvert av Togsettene.

## 4 Tilgang til data

Oppdragsgiver vil gjøre følgende data tilgjengelig for Leverandøren:

1. Beriket datasett, ref. punkt 3.3
3. Totalt antall Passasjerer som inkluderer faktiske tellinger og estimerte Passasjerer
4. Rapport på APC-utstyr som ikke leverer tellinger

Oppdragsgiver tilgjengeliggjør data via API og onlineportal.

Data gjøres tilgjengelig i det format som Oppdragsgiver bestemmer.

## 5 Endringer og justeringer i modellen

Alle operasjonelle justeringer i modellen kan gjennomføres av Oppdragsgiver uten at dette utløser Endringsordre etter reglene i Trafikkavtalens punkt 14. Slike justeringer er ikke begrenset til, men kan omfatte følgende:

- Bytte datakilder og benytte andre Sekundærdata fra andre kilder enn de som er benyttet i den initiale modellutviklingen.
- Justere vasking og prosessering av inndata jf. punkt 3.3 (Korreksjoner, vasking og beriking av Datagrunnlaget)
- Justere prosesseringen av data i modellen, jf. punkt 3.4.1 (Prosessering av data)
- Retrening og Kalibrering av modellen, jf. punkt 3.4.2 (Retrening og punkt 3.4.3(Kalibrering)

Oppdragsgiver skal til enhver tid holde Modelldokumentasjonen oppdatert slik at denne gjenspeiler de faktiske forhold.

Oppdragsgiver vil varsle leverandør med så fort det er besluttet at en endring skal gjennomføres. Endringer i modellens grunnprinsipper som er beskrevet i punkt 2.2 følger reglene om Endringsordre i Trafikkavtalens punkt 14.

Oppdragsgiver kan kreve endringer i Leverandørens forpliktelser til å levere informasjon (f.eks. å gi tilgang til nye og endrede datakilder for Telledata og/eller Produksjonsdata) dersom dette gir bedre presisjon. Leverandøren dekker sine egne kostnader, men kan kreve kompensasjon for dokumenterte merkostnader for å gi Oppdragsgiver tilgang til slike nye primærdatakilder dersom merkostnadene er direkte knyttet til Oppdragsgivers krav. Dersom merkostnadene utløses av krav fra andre tredjeparter, myndighetskrav eller Leverandørens egne behov dekkes dette av Leverandøren.



## Trafikkavtale Østlandet 2

- Regioneक्सpresstog (RE-tog) Skien - Eidsvoll
- Regioneक्सpresstog (RE-tog) Drammen - Lillehammer
- Regiontog (R-tog) Kongsberg - Eidsvoll
- Regiontog (R-tog) Drammen - Dal
- Regiontog (R-tog) Asker - Kongsvinger
- Regiontog (R-tog) Notodden - Porsgrunn (Bratsbergbanen)
- Regiontog (R-tog) Drammen - Oslo lufthavn
- Regiontog (R-tog) Stabekk/Oslo S - Oslo lufthavn

Vedlegg B-6 Statistisk modell [2.0](#)



Dokumentet er signert digitalt av:

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

Forseglet av



Posten Norge

Vedlegg B-6 Statistisk modell [2.0](#)

#### Endringsloggen

| Versjon             | Hva                                                                                                      | Dato                       |
|---------------------|----------------------------------------------------------------------------------------------------------|----------------------------|
| 1.0                 | Vedlegg B-6 av 23.01.2023.                                                                               | 28.06.2023                 |
| <a href="#">2.0</a> | <a href="#">Oppdatert tekst til punktene 1.1, 2.2, 3.1, 3.2.2, 3.3.1, 3.4.1, 3.4.2, 3.4.4, 3.5 og 5.</a> | <a href="#">05.11.2024</a> |

Dokumentet er signert digitalt av:

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

Forseglet av



Posten Norge

~~Underskrift og dato~~

~~Dette vedlegg er undertegnet i 2 – to –  
eksemplarer, hvorav Partene beholder ett  
eksempplar hver.~~

~~Oslo, xx.xx.2023~~

~~Oslo, xx.xx.2023~~

~~Knut Sletta~~

~~Gro Bakstad~~

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024



# Innhold

|                                                         |           |
|---------------------------------------------------------|-----------|
| <b>1 Innledning</b>                                     | <b>5</b>  |
| 1.1 Bakgrunn                                            | 5         |
| <b>2 Formål og grunnprinsipper</b>                      | <b>5</b>  |
| 2.1 Formålet med modellen                               | 5         |
| 2.2 Grunnprinsipper                                     | 5         |
| <b>3 Oversikt over prosessen</b>                        | <b>6</b>  |
| 3.1 Innledning                                          | 6         |
| 3.2 Datagrunnlaget (inputdata)                          | 6         |
| 3.2.1 Primærdata                                        | 6         |
| 3.2.2 Sekundærdata                                      | 7         |
| 3.3 Korreksjoner, vasking og beriking av Datagrunnlaget | 8         |
| 3.3.1 Vasking av Telledata                              | 8         |
| 3.3.2 Produksjonsdata                                   | 10        |
| 3.4 Behandling av data i statistisk modell              | 10        |
| 3.4.1 Prosessering av data                              | 10        |
| 3.4.2 Retrening                                         | 10        |
| 3.4.3 Kalibrering                                       | 11        |
| 3.4.4 Konduktørjustering                                | 11        |
| 3.4.5 Beregning av passasjertall i avvikssituasjon      | 11        |
| 3.5 Utdata fra statistisk modell                        | 12        |
| <b>4 Tilgang til data</b>                               | <b>12</b> |
| <b>5 Endringer og justeringer i modellen</b>            | <b>12</b> |



# 1 Innledning

## 1.1 Bakgrunn

Trafikkinntektene på linjer som driftes av Leverandøren skal fordeles gjennom en inntektsfordelingsmodell (jf. «Vedlegg B-7») som baserer seg på totalt antall Passasjerer fra de ulike togproduktene.

Oppdragsgiver har utviklet en statistisk modell som skal supplere tellinger med estimer for å sikre et best mulig bilde av totalt antall Passasjerer. Modellen skal estimere Passasjertall der Telldata mangler.

Den statistiske modellen er utviklet basert på historiske APC-tellinger, Produksjonsdata og Sekundærdata ~~fra 2018 og 2019~~. Utviklingen av modellen og datagrunnlaget som er benyttet i utviklingen er nærmere beskrevet i Modelldokumentasjonen.

Den statistiske modellen driftes og vedlikeholdes av Oppdragsgiver.

# 2 Formål og grunnprinsipper

## 2.1 Formålet med modellen

Den overordnede målsetningen med den statistiske modellen er å sikre et så presist som mulig og dermed rettferdig, transparent og troverdig passasjergrunnlag.

## 2.2 Grunnprinsipper

For å oppnå det overordnede formålet (i punkt 23.1) gjelder følgende grunnprinsipper for statistisk modell:

1. Statistisk modell utvikles for å gi mest mulig presise resultater til bruk i inntektsfordelingen
  - a. Med «resultater» menes modellens utregning av antall Passasjerer på en gitt linje
  - b. Med «mest mulig presis» menes det at modellens snittfeil er så lav som mulig i en måleperiode (per måned og per linje). Snittfeilen beregnes med Root Mean Square Error (RMSE), [Mean Biased Error \(MBE\)](#) og [Mean Average Error \(MAE\)](#) som [alle](#) er ~~en~~ statistiske metoder for å måle presisjonen i statistiske estimer.
2. Presisjonsnivået i estimatene kontrolleres månedlig mot faktiske tellinger fra APC-utstyr. [Togoperatør vil få tilgang til resultatene.](#)
3. Justeringer i statistisk modell implementeres kun dersom endringen ventes å gi lik eller forbedret presisjon
4. Planlagte justeringer i statistisk modell kommuniseres til Leverandøren dersom de ventes å ha en vesentlig effekt på resultatene



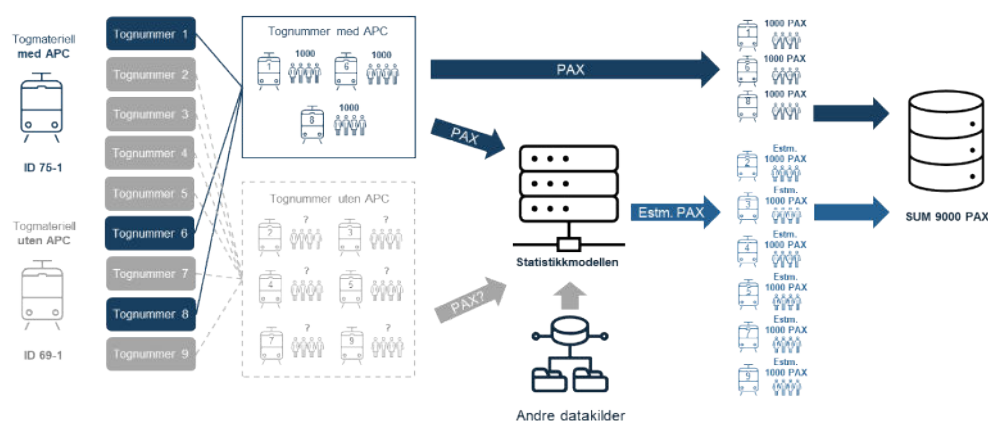
## 3 Oversikt over prosessen

### 3.1 Innledning

Proessen som Oppdragsgiver benytter for å estimere antall Passasjerer er i korte trekk følgende:

1. Oppdragsgiver innhenter datagrunnlaget som togoperatør tilgjengeliggjør (punkt 3.2)
2. Data vaskes og berikes (punkt 3.3)
3. Data behandles i statistisk modell ved at kunstig intelligens foreslår et regelsett for å estimere reiseaktivitet. Regelsettet implementeres i statistisk modell etter kvalitetssikring av Oppdragsgiver. [For utfyllende beskrivelse henvises det til Modelldokumentasjonen \(punkt 3.4\)](#) Proessen er illustrert i figuren under:

Figur 1: Illustrasjon av den statistiske modellen



### 3.2 Datagrunnlaget (inputdata)

Datagrunnlaget (Inputdata) som benyttes i den statistiske modellen består av primærdata (Telledata og Produksjonsdata) og Sekundærdata.

[Leverandørs leveranse av inputdata til Oppdragsgiver følger av Trafikkavtalenes bestemmelser avsnitt 6.7 \(levering og bruk av telledata og annen rådata\) og avsnitt 6.8 \(Rapporterings- og informasjonsplikt\) samt Trafikkavtalens vedlegg A - punkt 6.5 \(levering av telledata og annen rådata\)](#)

#### 3.2.1 Primærdata

Modellen benytter Telledata og Produksjonsdata som de sentrale inputkilder for å beskrive togtrafikken.

##### Telledata (fra APC-utstyr)

Telledata inneholder opplysninger om antallet på- og avstigninger på stasjonen, per dør, per Togsett, per stasjon, per tognummer. Basert på antall Påstigende og Avstigende, kan vi også utregne PAX på et gitt tog. APC-dataene utgjør dermed fundamentet for input til modellen.

### Produksjonsdata

Produksjonsdata gir informasjon om alle tog, også togene som ikke har APC-utstyr.

Produksjonsdata benyttes som grunnlag for hvor og når det kjører et tog, og hvilke tog som kjører. Dette er avgjørende for å estimere antall Passasjerer.

Produksjonsdataene gir også opplysninger om produksjonen, samt flere forskjellige variabler relatert til produksjonen, f.eks. forsinkelse, innstillinger, lokasjon, avgangstid, kapasitet osv.

### 3.2.2 Sekundærdata

Sekundærdata som benyttes kommer fra ulike tredjeparter. Datakildene som er benyttet i utviklingen av modellen er beskrevet i Modelldokumentasjonen. [Eksempler på sekundærdata er værdata, trafikkdata, kalenderinformasjon og områdedata/geografi.](#)

#### Andre sekundærdata/datakilder

Ved behov (eksempelvis ved bortfall av en datakilde) vil det gjøres vurderinger om det er nødvendig å benytte andre Sekundærdata enn de opprinnelig benyttet for å utvikle modellen, og hvilke datakilder som skal benyttes.

Sekundærdata kan hentes fra andre og/eller nye kilder dersom Oppdragsgiver finner det hensiktsmessig, f.eks. at eksisterende datakilder ikke er tilgjengelig.

#### Værdata

Værdata har som formål å beskrive hvordan været er på den gjeldende avgang og dag. Dette er med på å beskrive forskjeller i den aktuelle transportsituasjon, og flere studier viser at været påvirker transportatferd. Data dekker over temperatur, nedbør, vind og lufttrykk.

#### Trafikkdata

Trafikkdata og reisetider kan være en indikator (proxy) for flaskehalser i trafikken som smitter av på den kollektive trafikken. Det kan f.eks. være stengte linjer på ruten, antall stopp til nærmeste stengte stopp, samt om stoppet før eller etter på ruten er stengt.

#### Kalenderinformasjon

Kalenderdata er sentralt i forhold til å fange opp systematiske variasjoner, ved dataene. Kalenderinformasjon definerer typiske informasjoner om datoer som f.eks. helligdager, helger, ferier osv. Kalenderinformasjon kan også fange opp systematiske, daglige variasjoner som rushtid.

#### Områdedata og geografi

Modelleringen berikes ytterligere med del data om området for stasjonens plassering, f.eks. befolkningstetthet, antall stasjoner i området, antall pendlere, antall biler, og medianinntekt.



#### Andre Sekundærdata/datakilder

Ved behov (eksempelvis ved bortfall av en datakilde) vil det gjøres vurderinger om det er nødvendig å benytte andre Sekundærdata enn de som er beskrevet i punkt 3.2.2. og hvilke datakilder som skal benyttes.

Sekundærdata kan hentes fra andre og/eller nye kilder dersom Oppdragsgiver finner det hensiktsmessig, f.eks. at eksisterende datakilder ikke er tilgjengelig.

### 3.3 Korreksjoner, vasking og beriking av Datagrunnlaget

Før Telldata benyttes for beregning av antall Passasjerer, er det derfor nødvendig å vaske dataene for å forbedre datakvaliteten slik at antall Passasjerer blir beregnet mest mulig presist.

For å ivareta kvaliteten, gjennomføres følgende steg for å berike og vaske datasettet før det benyttes i statistisk modell:

1. Telldata mottas urørt fra APC-utstyr.
2. Telldata vaskes automatisk basert på et gitt regelsett.
3. Telldata og Produksjonsdata komplementeres med Sekundærdata.
4. Kontrollmekanisme validerer tallene (for eksempel ekstremverdier).
5. Det vaskede og berikede datasettet benyttes i modellen for estimering.

For utfyllende beskrivelsen av punktene henvises det til modelldokumentasjon.

#### 3.3.1 Vasking av Telldata

Utjevning av antall Passasjerer

En grundig gjennomgang av Telldata synliggjør utfordringer med datasettet som må løses. Dette skyldes to forhold:

- 1) Målefeil: All sensordata vil alltid inneholde en viss grad av feil – dvs. at systemene noen ganger teller en påstigende passasjer i stedet for to. Dersom systemet teller antall avstigende på korrekt måte betyr dette at vi vil ha negativ PAX på toget siste stopp.
- 2) Kunstig ulogisk oppførsel fra de reisende: Eksempelvis forventes det ikke avstigninger på første stopp, men dette forekomme eksempelvis hvis man har hjulpet noen inn på toget. På- og avstigninger fra ombordpersonalet er en annen kilde til slike feil.

(Målefeil og) ulogiskheter i av- og påstigningsmønster forventes å være tilfeldig på tvers av tog. Målefeilen vil allikevel sannsynligvis være korrelert med ytre faktorer som f.eks. passasjertall. Dermed kan målefeilen være større i rushtid enn utenfor rushtid og på større stasjoner sammenlignet med mindre stasjoner.

Begge disse typene feil håndteres ved at de internaliseres i dataene. Dette vil si at dataene endres ved at de må oppfylle et sett med regler. Dette fører til en utjevning av antall passasjerer over togets reise.

En sentral operasjon er nullstilling av togene slik at antall Passasjerer på toget ikke er negativ ved slutten av reisen eller under reisen eller at det ikke er Passasjerer som aldri går av.



Denne nullstillingen gjøres i følgende rekkefølge; først gjøres det opp differanse ved endt tur, for å så fordele differansen ut over hele turen proporsjonalt med antallet av telte Passasjerer ved stoppet.

Dette sikrer at hver tur slutter med null Passasjerer. Har vi et negativt antall Passasjerer ombord underveis, utlignes dette ved å tilføye Påstigende og Avstigende Passasjerer. [For utfyllende beskrivelse av prosessen henvises det til Modelldokumentasjonen avsnitt 1.0.](#)

Hvis det ikke er mulig å utjevne hensiktsmessig, betrakter vi tellingene som utroverdige og fjerner disse dataene fra utvalget som benyttes i modellen.

[Overordnet er det viktig at målefeil på Telleutstyret er tilfeldige og videre forutsetninger er eksempelvis at:](#)

- ~~Telleutstyret har like begrensninger med tanke på å telle mange Passasjerer, eller få Passasjerer. Med andre ord antas det at evt. målefeil ikke varierer med antall Påstigende.~~
- ~~Det er noe systematikk i hvor, og når Telleutstyret feiler, slik at vi ikke får laget en registrering.~~
- ~~Telleutstyret teller barn, hunder, rullestoler, sykler eller lignende på en ensartet måte, og f.eks. en sykkel gjør ikke at personer telles dobbelt.~~
- ~~Presisjonen til Telleutstyret avhenger ikke av årstid, tid på dagen eller lignende variasjoner. Vi har lagt til grunn at yttertøy, eller koffert og annet reiseutstyr ikke fører til at personer blir talt dobbelt.~~
- ~~Vaskeprosedyren på Telleutstyret før vi får dataene er ikke systematiske, og vasker særlige tellinger bort~~
- ~~Telleutstyret er presist og teller ikke skjevt~~

Dersom det oppdages endringer i forutsetningene, vil metodikken og reglene for utjevning endres.

#### Korrigerende og forenklinger av Telledata

Telledata som inneholder observasjoner som ikke oppfyller de logiske regler som er beskrevet nedenfor fjernes.

1. **For å unngå dobbelttelling kan en observasjon kun inngå én gang i modellen.** Derfor fjernes reiser hvor det er flere APC-tellinger på samme stopp på samme avgang. For å unngå dobbelttelling kan en observasjon kun inngå én gang i modellen.
2. **Avganger som er innstilt helt fjernes fullstendig.** Dette ettersom vi da vet at det ikke eksisterer noen Passasjerer. Reiser hvor enkelte stopp er innstilt beholdes, ettersom vi ellers ville vært nødt til å fjerne store mengder informative data. Det innstilte stoppet fjernes siden vi her vet at ingen går hverken på eller av.

**Reiser som ikke er komplette fjernes.** Det vil si at en reise må ha et første og siste stopp, samt alle stoppene imellom (bortsett fra innstilte stopp). For disse ufullstendige reisene er det ikke mulig å avgjøre hvorfor et stopp mangler. Derfor kan vi ikke vite hvor mange av- og påstigninger det er på stoppet. For å sikre kvalitet i modellen, blir disse reisene frasortert under trening av modellen, men blir tatt tilbake når modellen benyttes til prediksjon av Passasjertall i output. Dette for å sikre høyest mulig dekningsgrad.

[For avganger med togsett med APC-utstyr hvor det er mottatt APC-tall på deler av strekningen, men ikke for alle stasjoner, vil det Jernbanedirektoratet løpende vurderes og mulig implementeres tiltak for bruk av tilgjengelig APC. Slike tiltak vil kun implementeres om de](#)



[vurderes til å heve presisjonen i modellen og forutsettes at det er mulig å ha tilstrekkelig gode kvalitetssikringsmekanismer, som unngår at støy fra APC skaper vesentlige utslag.de-s](#)

3. **Vi samler det gitte togets dører i par.** Vi vet ikke på hvilken side av toget døren er åpen. Ved å samle dørene i par, gjør vi datasettet halvparten så stort. [Dette muliggjør en raskere estimering, og dermed mer effektive modeller uten at informasjon blir forkastet.](#)
4. **Vi vet ikke hvilken vei toget kjører.** Dette betyr at vi ikke vet om dør 1 og 2 er de første 2 dørene som kjører inn på stasjonen, eller de siste. For å ikke skape unødige mye støy i modellen, koder vi dørene sammen, slik at vi identifiserer de fremste og siste dørpar som ender, og de midterste dørnumrene som midtdører.
5. **Passasjerer kan ikke bevege seg mellom to Togsett.** I utregningen av på- og Avstigende per Tog-leg antas det imidlertid at Passasjerene ikke kan bevege seg mellom de forskjellige Togsettene, og analysen foretas derfor på Togsettnivå. For noen avganger forekommer det derfor flere Togsett (vehicles) i samme tog.

### 3.3.2 Produksjonsdata

Før Produksjonsdata benyttes, kan det i enkelte tilfeller være nødvendig å fylle ut enkelte hull i dataene (f.eks. kan dette være manglende tidsangivelser eller kjøretøytype). For deretter å matche dette med andre kilder.

Følgende regler benyttes for behandling av Produksjonsdata som grunnlag for å utfylle enkelte hull i dataene. Disse prosedyrene sikrer at dataene har nødvendig kvalitet, før de blir brukt til estimering.

- I dataene forekommer det at ankomsttiden er etter avgangstiden, og at det er manglende tidsangivelser. Vi fyller hullene ut ved å estimere kjøretidene mellom stasjonene. Hvis det er en sammenheng med de manglende tidsangivelsene og for eksempel forsinkelse, så vil vi kunne underestimere antallet forsinkede tog, som igjen vil kunne påvirke modellen.
- For å finne antall dører og kjøretøy for avganger, må man vite hvilken kjøretøytype og hvor mange kjøretøy som håndterer avgangen. Datakilden for dette er data om togsammensetning som inneholder Dato og Tognummer, samt modell og antall kjøretøy. I en del tilfeller vil man ikke få treff i denne listen pga. hull i datagrunnlaget. På bakgrunn av datakilden for togsammensetning, er det derfor laget en mapping basert på hvilke togsammensetninger som er vanligst per TogNummer (uten dato). Denne brukes så for å avgjøre modell og antall kjøretøy for avganger som mangler faktisk avgang i datakilden for togsammensetning. I de tilfeller hvor det fortsatt ikke finnes togsammensetningsdata på tognummer-nivå, brukes tilsvarende mapping på linjenivå

## 3.4 Behandling av data i statistisk modell

### 3.4.1 Prosessering av data

All Produksjonsdata og Sekundærdata prosesseres for å lage et sett variabler som er enkle for den statistiske modellen å jobbe med. Utformingen av disse variablene kalles "Feature Engineering", og beskrives nærmere i Modelldokumentasjonen. [I denne prosessen kan det også bestemmes at enkelte datafelter ikke benyttes fordi de ikke bidrar til økt presisjon. Utformingen kan endres over tid, men gjøres alltid med formål om å maksimere modellens presisjon.](#)

### 3.4.2 Retrening

[Modellen har benyttet datagrunnlag fra 2018 og 2019 som grunnlag for treningen av modellen.](#)



En Retrening vil si at regelsettet oppdateres i lys av nytt datagrunnlag samlet inn siden siste Retrening. Det betyr at en Retrening eksempelvis kan bety at modellen oppdateres for å bedre beskrive nye reisemønster som har blitt observert i befolkningen. Dette gjøres for å sikre at modellen fortsatt predikerer på best mulig måte. Hvor ofte modellen Retrenes i fremtiden vil avhenge av (a) endring i presisjonsnivå i estimeringene og (b) konsekvensene av ytre faktorer som f.eks. store endringer i togtilbud, større kommunesammenslåinger og pandemier.

Hvor ofte modellen Retrenes i fremtiden vil avhenge av (a) endring i presisjonsnivå i estimeringene og (b) konsekvensene av ytre faktorer som f.eks. store endringer i togtilbud, større kommunesammenslåinger og pandemier.

### 3.4.3 Kalibrering

Kalibrering av den statistiske modellen kan i ekstraordinære tilfeller være nødvendig for å sikre at estimatene har riktig størrelse, men hvor man ikke har tilstrekkelig data til å Retrene modellen. Dette er spesielt relevant når det oppstår strukturelle endringer i trafikkmønsteret på kort tid (f.eks. etter Covid-19) som gjør at treningsdataene i en periode ikke er representative for nytt trafikkmønster. I slike tilfeller vil en kortsiktig Kalibrering gi forbedret presisjon mens det samles historikk for å gjennomføre en Retrening. Ved en eventuell kalibrering vil Oppdragsgiver gi Leverandør innsyn i beslutningsgrunnlaget. For å sikre at estimatene er så riktig som mulig er det innført en teknikk i modellen som muliggjør kalibrering av estimerte passasjertall mot gyldige APC-tall. APC-tall er den mest tilgjengelige og pålitelige kilden til informasjon, og derfor benyttes dette datagrunnlaget i kalibrering av estimatene.

Kalibreringen har vist seg å føre til betydelig reduksjon i skjevheten til estimatene. Ved å innføre denne metoden har risikoen for at nye modeller fører til et nivåskifte i passasjertall blitt betydelig redusert.

Modellen vil kun kalibrere ved behov.

### 3.4.4 Konduktørjustering

For å hindre at konduktørens på- og avstigninger bidrar til kunstig høye passasjertall vil det gjøres en nedjustering av antall påstigninger og avstigninger i et forsøk på å minimere effekten av konduktørens aktivitet. Det er usikkerhet knyttet til akkurat hvor mange påstigninger og avstigninger som stammer fra konduktøren. Konduktørjusteringen i Statistisk modell baserer seg derfor på en forutsetning om at én konduktør per togsett går på første stasjon, av på siste stasjon og av og på per stopp imellom. Justeringen skal utligne disse, i den grad justeringen ikke fører til negative verdier for påstigninger, avstigninger og antall passasjerer på toget mellom to stopp.

### 3.4.5 Beregning av passasjertall i avvikssituasjon

Statistisk modell beregner passasjertall også ved bruk avvikstransport (buss for tog).

-er

- Buss for tog estimeres som om det var opprinnelig avgang med tog som gikk
- Dersom det går buss for tog i parallell med tog, vil det ikke estimeres passasjertall for bussen
- Statistisk modell er tilpasset slik at man kan benytte APC-tall der hvor dette eksisterer for en delavgang med tog

Modellen estimerer også passasjertall ved avvikstransport hvor de reisende blir henvist til andre kollektivaktørers transportmidler – eksempelvis T-banen i Ruter-området. Disse tallene brukes dog kun til inntektsfordeling mellom aktørene Østlandet 1 og Østlandet 2 og vil ikke inngå i totalt antall passasjerer på jernbanen. Passasjerkilometer som produseres på kollektivaktørene sine transportmidler ved avvik kommer til fratrekk i oppgjøret i takstsamarbeidsavtalen.



### 3.5 Utdata fra statistisk modell

Følgende nøkkeltall er output fra statistisk modell:

1. Antall Passasjerer om bord i toget etter passasjerutveksling.
2. Brutto Påstigende per tog.
3. Netto passasjerutveksling per tog.
4. Brutto Avstigende per tog.
5. Netto passasjerutveksling pr. dør.

Netto passasjerutveksling pr dør beregnes ut fra en fordelingsnøkkel. Fordelingsnøkkelen er basert på historiske Telledata for hvordan Passasjerer fordeler seg mellom dørene i tog. Da fordelingsmønstrene varierer mellom togtyper og stasjoner, beregnes fordelingsnøkkelen for hver enkel kombinasjon av Tog-leg og Stasjon.

Modelldokumentasjonen [punkt 6](#) beskriver hvordan statistisk modell benyttes for å beregne nøkkeltallene.

Modellen benytter to separate [maskinlæringsmodeller](#) [Random Forest modeller](#) for å beregne Passasjerer på togene. Den ene modellen estimerer antall Påstigende. Den andre modellen estimerer antall Passasjerer. Antall Påstigende og Passasjerer i toget estimeres på Togsettnivå. For tog som består av flere Togsett, beregnes antall Passasjerer i toget som summen av estimatene i hvert av Togsettene.

## 4 Tilgang til data

Oppdragsgiver vil gjøre følgende data tilgjengelig for Leverandøren:

1. Beriket datasett, ref. punkt 3.3
3. Totalt antall Passasjerer som inkluderer faktiske tellinger og estimerte Passasjerer
4. Rapport på APC-utstyr som ikke leverer tellinger

Oppdragsgiver tilgjengeliggjør data via API og onlineportal.

Data gjøres tilgjengelig i det format som Oppdragsgiver bestemmer.

## 5 Endringer og justeringer i modellen

Alle operasjonelle justeringer i modellen kan gjennomføres av Oppdragsgiver uten at dette utløser Endringsordre etter reglene i Trafikkavtalens punkt 14. Slike justeringer er ikke begrenset til, men kan omfatte følgende:

- Bytte datakilder og benytte andre Sekundærdata fra andre kilder enn de som er [benyttet i den initiale modellutviklingen, beskrevet i punkt 3.2.2 \(Sekundærdata\)](#)
- Justere vasking og prosessering av inndata jf. punkt 3.3 (Korreksjoner, vasking og beriking av Datagrunnlaget)
- Justere prosesseringen av data i modellen, jf. punkt 3.4.1 (Prosessering av data)



- Retrening og Kalibrering av modellen, jf. punkt 3.4.2 (Retrening [og punkt 3.4.3\(Kalibrering\)](#)).

Oppdragsgiver skal til enhver tid holde Modelldokumentasjonen oppdatert slik at denne gjenspeiler de faktiske forhold.

[Oppdragsgiver vil varsle leverandør med så fort det er besluttet at en endring skal gjennomføres.](#)

Endringer i modellens grunnprinsipper som er beskrevet i punkt 2.2 følger reglene om Endringsordre i Trafikkavtalens punkt 14.

Oppdragsgiver kan kreve endringer i Leverandørens forpliktelser til å levere informasjon (f.eks. å gi tilgang til nye og endrede datakilder for Telledata og/eller Produksjonsdata) dersom dette gir bedre presisjon. Leverandøren dekker sine egne kostnader, men kan kreve kompensasjon for dokumenterte merkostnader for å gi Oppdragsgiver tilgang til slike nye primærdatakilder dersom merkostnadene er direkte knyttet til Oppdragsgivers krav. Dersom merkostnadene utløses av krav fra andre tredjeparter, myndighetskrav eller Leverandørens egne behov dekkes dette av Leverandøren.

- MARGARETH NORDBY-KRINGLI, 20.12.2024
- ERIK LUND, 20.12.2024

