

Mot aktivitetsbasert etterspørselsmodellering

Bruk av maskinlæring for prediksjon av transportmiddelvalg og aktivitetsplaner

TØI rapport 2065/2024 • Forfattere: Stefan Flügel, Christian Weber, Simen Sørbøe Klommestein, Johan Korsmo, Anders Kielland • Oslo, 2024 • 62 sider

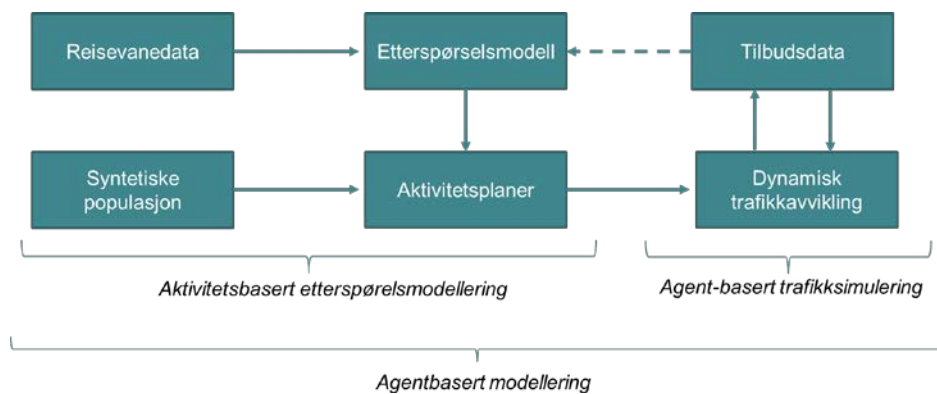
Denne rapporten dokumenterer FoU-arbeidet som ble utført som en del av PRELONG-prosjektet mellom 2022 og 2024, med fokus på å generere aktivitetsplaner for en syntetisk befolkning i Stor-Oslo-området. Kjernen i tilnærmingen vår er to nevralt nettverksmodeller som forutsier egenskapene til alle reiser som foretas på en typisk ukedag. Disse maskinlæringsmodellene er trent på reisevanedata fra RUTER-MIS (2017–2024) og anvendes på befolknings- og pendlerdata.

Transportmodellering utvikler seg for å møte utfordringene knyttet til urbanisering, teknologiske fremskritt og komplekse reisemønstre. Tradisjonelle firetrinnsmodeller for transportetterspørsel, selv om de er grunnleggende, klarer ofte ikke å fange opp den dynamiske naturen til reisemønstre i urbane områder.

Aktivitetsbaserte reiseetterspørselsmodeller (ABDM) gir en mer detaljert og dynamisk representasjon av reisemønstre, som tar hensyn til kompleksiteten i daglige tidsplaner, individuelle preferanser og miljømessige begrensninger. Dette oppnås ved: 1) Den aggregerte datastrukturen erstattes med en disaggregeret struktur, dvs. individer (i stedet for befolkningssegmenter innenfor geografiske områder) er analyseenheten. 2) Transportetterspørsel er avledet fra etterspørselen etter å utføre aktiviteter, med fokus på sekvenser av turer eller hele-dagen-reiseplaner i stedet for individuelle (tur-retur) reiser. 3) Tid (klokkeslett) modelleres eksplisitt (start-/sluttidspunkter og aktivitetsplanlegging er viktige).

For å bygge bro mot aktivitetsbaserte og agentbaserte transportmodeller dokumenterer denne rapporten genereringen av aktivitetsplaner (først dokumentert i Flügel 2022) ved bruk av en maskinlæringstilnærming, spesielt dype nevralt nettverk.

Figur S1 gir en skjematisk oversikt over hoveddataene, metodene og konseptene, og hvordan disse henger sammen i vår overordnede tilnærming til transportmodellering. Denne rapporten beskriver de fire rutene på venstre side av figuren, med spesielt fokus på etterspørselsmodellen, som i vårt tilfelle består av to nevralt nettverksmodeller (modell 1 og modell 2).



Figur S1: Oversikt over relaterte data og metoder i denne rapporten.

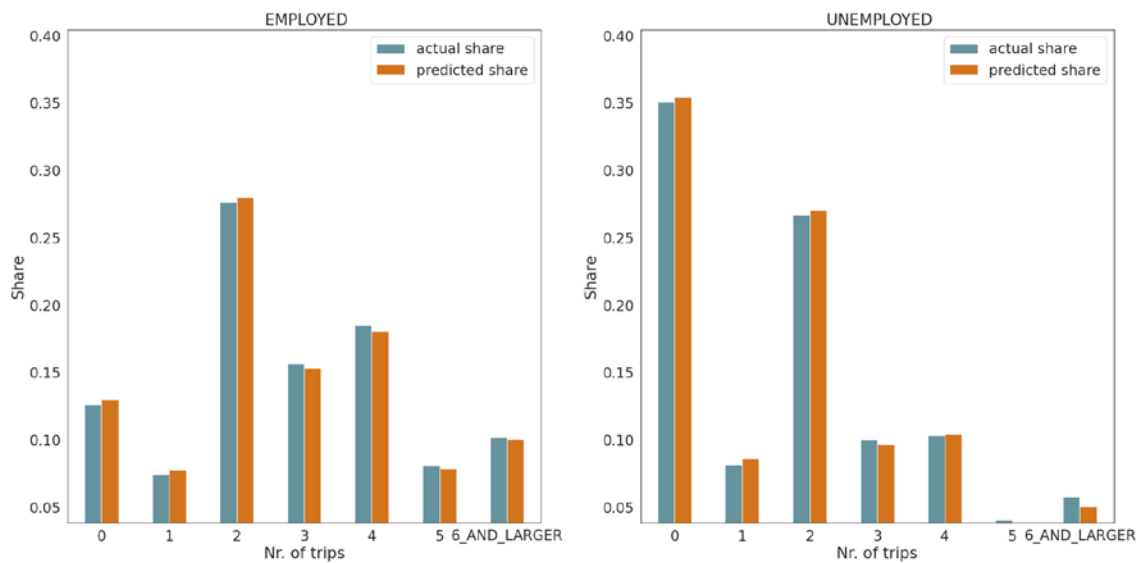
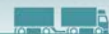
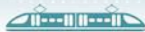
Som illustrert i Figur S2 forutsier modell 1 de overordnede egenskapene til reisedagboken (aktivitetsplaner): Starttidspunkt for første reise, starttidspunkt for siste reise, antall reiser og rekkefølgen av reiser før og etter arbeidsaktiviteten. Modell 2 forutsier egenskaper ved reisene: Reisemåte, formål, endesone og tidsvarighet siden forrige reise.



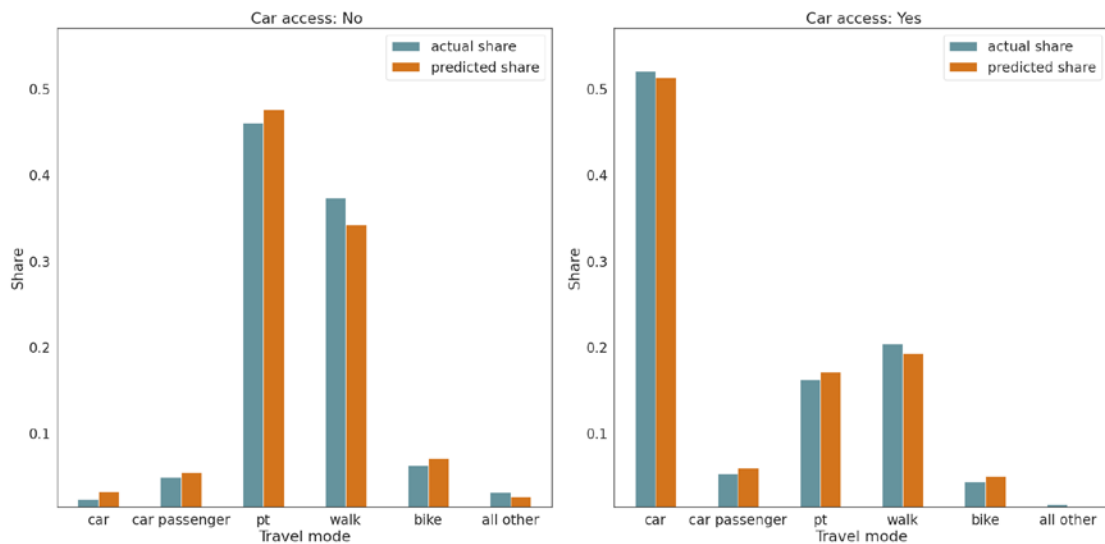
Figur S2: Inndata- og utdata-variabler for de to nevrale nettverksmodellene. De lilla klammene indikerer modell 1, og den grønne klammen indikerer modell 2. Kategoriske variabler er i **fet** skrift.

De stiplede pilene indikerer at utdata blir transformert til inndata når modellene brukes på nye data. Vår tilnærming fanger opp korrelasjoner mellom flere reiseegenskaper (reisehensikt, transportmiddel, avreisetidspunkt, destinasjonsvalg). For eksempel vil «hensikt: hente» være sterkt korrelert med resultatet «transportmiddel: bil».

De nevrale nettverksmodellene er trent på et tilfeldig utvalg av 75 % av respondentene i Ruter-MIS fra januar 2017 til september 2024. Modellene er validert mot de øvrige 25 % av utvalget (valideringsdata). Samlet sett ser modellene ut til å fange korrelasjonsmønstrene i valideringsdataene, som vist i de neste to figurene.



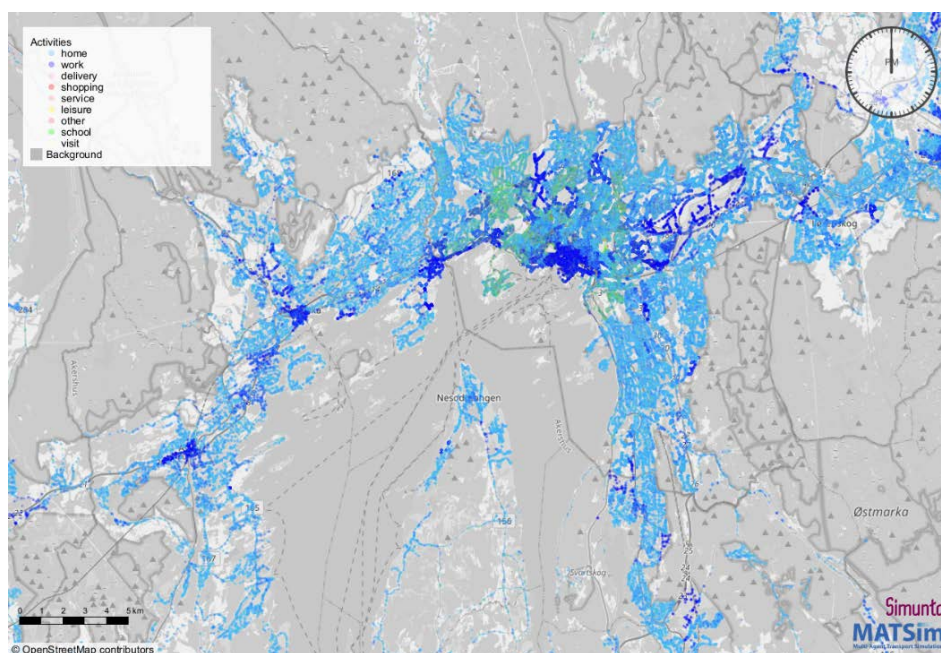
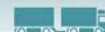
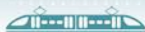
Figur S3: Faktisk andel (i Ruter-MIS) og forutsagt andel (fra modell 1) på valideringsdataene for antall turer gitt yrkesstatus.



Figur S4: Faktisk andel (i Ruter-MIS) og forutsagt andel (fra modell 1) på valideringsdataene for transportmiddelvalg gitt biltilgang.

Modellene er så anvendt på en syntetisk befolkning som er sammensatt av forskjellige data kilder (registerdata, pendlerstatistikk og sonedata fra transportmodeller).

Figur S5 viser den geografiske fordelingen av ulike aktiviteter klokka 12.



Figur S5: Geografisk fordeling av genererte aktiviteter ved klokkeslett 12:00:00.

Rapporten inkluderer noen tilleggsanalyser om effektene av Level-of-Service på prediksjonen av transportmiddelvalg. Vi finner at etterspørselastisitetene, slik de er implisert av prediksjoner av de nevrale nettverksmodeller, avhenger sterkt av modellformuleringen og er generelt noe lave.

Vår overordnede tilnærming har flere kjente begrensninger som krever videre utvikling. For det første fører det grove sonesystemet som brukes til destinasjonsprediksjon til upresise destinasjoner for ikke-arbeidsreiser, noe som potensielt kan resultere i ulogiske eller altfor lange turer. Fremtidige versjoner bør inkludere prediktive modeller for reiselengde, geografisk betinget utvalg eller grunnkrets-spesifikke destinasjonsvalgmodeller for å adressere disse problemene. I tillegg kan populasjonssyntese dra nytte av rikere datainnganger, som mikrodata.no, og bruk av maskinlæring eller generativ AI for å forbedre agentkarakteristikker som husholdstype, utdanning og inntekt.

Til tross for forbedringer siden tidligere versjoner genererer modellen fortsatt for mange turer på ettermiddagen, sannsynligvis på grunn av prediksjonsmetoden som predikerer tur-for-tur. Utforskning av avanserte arkitekturer som rekursive nevrale nettverk kan muligens bidra til å løse dette problemet. Validering av syntetiske aktivitetsplaner forblir utfordrende på grunn av begrensede eksterne data, der mobildata eller MATSim-validerte trafikksimuleringer kan tilby potensielle løsninger. Til slutt er den nåværende backend-implementasjonen, som er avhengig av Python-skript og notatbøker, lite brukervennlig. En front-end-grensesnitt er planlagt for 2025 for å forbedre funksjonaliteten og lette scenariotesting.

Til tross for disse svakheter viser tilnærmingen betydelig potensial. Dens kompatibilitet med dynamiske transportmodeller og anvendelighet i agentbasert modellering muliggjør videre anvendelser, som transportsimuleringer i MATSim. Videre har den romlig-temporale datastrukturen potensiale utover transport, for eksempel i epidemiologiske simuleringer.

Metodens anvendbarhet til fremtidige syntetiske populasjoner (f.eks. for 2030 eller 2050) sikrer relevans for langsiktige prediksjoner, forutsatt at inngangsdata forblir representative. Tilnærmingen kan bygge bro mellom fremtidige agentbaserte arealbruksmodeller som forutsetter langsiktig atferd og transportsimuleringsmodeller som fanger opp detaljerte atferdsjusteringer i transportnettverket. Dette markerer et viktig steg mot å fremme avanserte transportmodeller i norske byer.