

Demopan-2
Att kombinera enkätdata och
mobilnätdata för bättre skattning
av färdmedelsvalmodell för
långväga resor

The logo for VTI (Västra Transportinstitutet) consists of the lowercase letters 'vti' in a bold, black, sans-serif font. A thin red vertical line is positioned to the left of the logo.

Populärvetenskaplig sammanfattning

Ida Kristoffersson (VTI)
Angelica Andersson (VTI)
Maria Börjesson (VTI)

VTI opublicerad avrapportering

Demopan-2
**Att kombinera enkätdata och
mobilnätsdata för bättre skattning av
färdmedelsvalmodell för långväga resor**
Populärvetenskaplig sammanfattning

Ida Kristoffersson (VTI)

Angelica Andersson (VTI)

Maria Börjesson (VTI)

Författare: Ida Kristoffersson VTI, Angelica Andersson VTI, Maria Börjesson VTI.

Diarienummer: 2022/0042-7.1

Denna avrapportering ingår inte i någon av VTI:s publikationsserier
och publiceras därmed inte av VTI.

Datum: 2023-12-20

Innehållsförteckning

| | |
|---|-----------|
| Godkänd avrapportering | 6 |
| Sammanfattning | 7 |
| 1. Bakgrund | 8 |
| 2. Syfte och mål | 9 |
| 3. Resultat 1: Att kombinera mobilnät- och resvanedata för skattning av logit-modell för långväga färdmedelsval | 10 |
| 3.1. Bakgrund och frågeställning | 10 |
| 3.2. Metod | 10 |
| 3.3. Resultat och lärdomar | 11 |
| 4. Resultat 2: Att tillämpa maskininlärningsmetoder vid skattning av färdmedelsvalmodell på kombinerade datakällor | 12 |
| 4.1. Bakgrund och frågeställning | 12 |
| 4.2. Metod | 12 |
| 4.3. Resultat och lärdomar | 13 |
| 5. Resultatspridning | 14 |
| Referenser | 15 |

Godkänd avrapportering

Denna avrapportering har genomgått peer review den *11 december 2023* och ansvarig chef *Magnus Berglund* har den *20 december 2023* godkänt denna avrapportering för leverans till uppdragsgivare.

Sammanfattning

Denna rapport beskriver arbetet i projektet Demopan-2 med att utveckla prognosmodeller för färdmedelsval för långväga resor givet att resvaneundersökningar har sjunkande svarsfrekvenser, medan mobilnätsdata framkommit som en potentiell datakälla för resebeteende. I Demopan-1 skattade vi färdmedelsvalsmodeller med mobilnätsdata som enda källa till data om resebeteende. Vi såg då problem i skattningen av modeller på mobilnätsdata. Det ena problemet uppstår på grund av att mobilnätsdata inte innehåller information om ärende (till exempel om resan är en privat resa eller en tjänsteresa). Det andra problemet för att det är svårt att särskilja bil och buss genom automatisk identifikation av färdmedel eftersom de båda använder väginfrastrukturen. I Demopan-2 kombinerar vi mobilnätsdata med enkätdata, för att lösa dessa problem. Vi skattar en klassisk logit-modell genom att använda båda datakällorna och finner att detta förbättrar modellskattningarna när det gäller beteendemässig rimlighet. Genom att genomföra en litteraturöversikt över maskininlärnings-metoder (ML-metoder) och deras tillämpning på färdmedelsval undersöker vi också potentialen för helt nya typer av modeller som skulle kunna ersätta den klassiska logit-modellen. Här finner vi att ML har potential vid kombination av mobilnätsdata och enkätdata, dels eftersom det finns tidigare använda metoder inom ML-området för att hantera att den ena datakällan har högre upplösning, dels eftersom man inte behöver ansätta i förväg vilka variabler som förklarar indelningen i olika ärende-klasser utan detta kan göras data-drivet.

1. Bakgrund

Projektet Demopan-2 finansieras av Trafikverket (TRV 2021/22422) och pågår från februari 2022 till december 2023. Det är ett uppföljningsprojekt till projektet Demopan-1 och utgör andra halvan av Angelica Anderssons doktorandarbete. Projektet finansieras av Trafikverket och projektledare är Ida Kristoffersson (även bi-handledare till Angelica). Projektmedverkande är Angelica Andersson och Maria Börjesson (även huvudhandledare till Angelica). I juni 2022 tog Angelica sin Licentiat-examen. Licentiat-avhandlingen (Andersson, 2022) blir således en leverans från projektet även om den inte står med som leverans i projektspecifikationen.

Svarsfrekvensen på resvaneundersökningar har sjunkit drastiskt de senaste decennierna. För att kunna skatta modeller som kan användas för prognoser på färskdata är det mycket troligt att nya datakällor kommer behöva användas. Passivt insamlade data så som mobilnätsdata ger stora datamängder. De kan ersätta eller komplettera resvaneundersökningar för att man kan se vilken mobilnätsantenn varje telefon med ett visst abonnemang har kopplats till. Utifrån antennernas registrering och position kan man sedan rekonstruera start- och målpunkter för de resor som telefonen har varit med på och identifiera färdmedel genom att jämföra antennsignaler på olika platser med koordinaterna för väg- och järnvägsinfrastruktur samt flygplatser. Mobilnätsoperatören har skäligen anledning att registrera vilka antenner mobiltelefonerna har kopplats till för att kunna bedriva sin verksamhet, och endast aggregerad data utan individuella abonnemangsnummer tillåts lämna mobiloperatörens servrar av integritetsskäl. En nackdel är dock att man inte har information om ärende eller resenärens socioekonomi kopplad till varje resa. Det är också svårt att i mobilnätsdata skilja på olika färdmedel som rör sig på väg, eftersom färdmedelsklassificeringen i stor utsträckning baseras på koppling till infrastrukturen.

I Demopan-1 skattade vi färdmedelsvalmodeller för långväga resor baserade på endast mobilnätsdata. Modeller för långväga resor passar bra för analys av mobilnätsdata eftersom zonerna i långväga modellen är relativt stora. Den geografiska precisionen i vart mobiltelefonen befunnit sig vid början och slutet av en resa behöver därmed inte vara lika hög som för regionala resor med finare zonindelning. Resultaten från Demopan-1 visade att det gick att skatta färdmedelsvalmodeller för långväga resor på endast mobilnätsdata. En nästlad logitmodell användes för att dela upp väg i färdmedlen buss och bil baserat på skillnader i utbud. För att klassificera resor på olika ärenden (privata respektive tjänsteresor) användes en latent-klass-modell. Modellens tidsvärden blev dock mycket höga för tjänsteresor och för bussresor, vilket tyder på att modellen inte lyckas fånga uppdelningen både i ärende (tjänste/privat) och färdmedel på väg (buss/bil) så bra.

2. Syfte och mål

Syftet med Demopan-2 är att kombinera enkätdata och mobilnätsdata för att se om tillägget av enkätdata kan förbättra skattningen av en färdmedelsvalmodell för långväga resor. Till exempel förbättra hur väl modellen lyckas modellera ärende och uppdelningen av väg i buss och bil. Syftet är också att utveckla och beskriva metoden för hur enkätdata och mobilnätsdata kan kombineras i skattning av långväga färdmedelsvalmodell.

I ett första steg används samma flernivå-latenta-klass-logit-ansats som i Demopan-1, med skillnaden att enkätdata läggs till mobilnätsdata och ett antal variabler som bara finns i enkätdata (sällskapsstorlek, körkortsinnehav m.m.) adderas. I ett andra steg testas en helt ny ansats baserad på en maskininlärningsmetod och potentialen av den nya metoden utvärderas. Syftet med detta andra steg är att se om nya typer av data också behöver nya metodansatser för att skatta en så bra prognosmodell som möjligt.

Syftet med Demopan-2 som beskrivits ovan kondenseras till två mål för projektet:

- 1) Att genomföra en skattning av en färdmedelsvalmodell för långväga resor i Sverige som använder både mobilnätsdata och resvanedata i skattningsprocessen och nyttjar respektive datakällas fördelar.
- 2) Att undersöka mer avancerade modeller/modellspecifikationer och utvärdera deras potential.

3. Resultat 1: Att kombinera mobilnätets- och resvanedata för skattning av logit-modell för långväga färdmedelsval

Angelica Andersson, Ida Kristoffersson, Andrew Daly, Maria Börjesson

Manuskript under granskning: Andersson et al. (2023, Preprint)

3.1. Bakgrund och frågeställning

Traditionellt har man använt enkätdata från resvaneundersökningar (RVUer) för att skatta parametrar till färdmedelsvalsmodeller. Modellerna används inom ett modellsystem för att prognosticera framtida reseefterfrågan under antaganden om en viss ekonomisk utveckling och för olika policyscenarier; till exempel byggandet av ny infrastruktur, förändringar i resekostnader genom skatter eller subventioner mm. Den prognosticerade reseefterfrågan är en nyckelkomponent i samhällsekonomiska nyttan (konsumentöverskottet), som behövs för att göra samhällsekonomiska nytto-kostnads-kalkyler. Den prognosticerade reseefterfrågan med olika färdmedel kan också användas för att prognosticera förväntade utsläppsökningar, eller utsläppsminskningar till följd av olika policybeslut. Eftersom svarsfrekvensen på resvaneundersökningarna har sjunkit under många år minskar också tillförlitligheten hos prognosmodellerna eftersom det inte längre kan antas att de preferenser som härleds ut data är representativa för hela befolkningen. Det har bland annat påpekats att personer som har fullt upp är mindre benägna att svara på enkäter (Stopher and Greaves, 2007). Det innebär ett problem för modellskattningen eftersom dessa personer troligen hade gjort andra avvägningar mellan restid och reskostnad än de personer som man lyckats få med i enkäten, vilket leder till att parametrarna i modellen inte är representativa för resenärerna och därför sannolikt leder till felaktiga effektprognoser. Problemet med sjunkande svarsfrekvenser är särskilt påtagligt för långväga resor, som utförs sällan jämfört med regionala resor och därmed är svårare att fånga på ett representativt sätt i RVUer. I det föregående forskningsprojektet Demopan-1 har färdmedelsvalsmodeller framgångsrikt skattats på enbart mobilnätetsdata i stället för enkätdata, men tidsvärden för buss och tjänsteresor blev klart högre än förväntat, vilket tyder på att det kan finnas fördelar med att kombinera enkätdata med mobilnätetsdata, snarare än att direkt ersätta enkätdata med mobilnätetsdata. Därför kombinerar vi i denna artikel enkätdata och mobilnätetsdata för att skatta långväga färdmedelsvalsmodeller.

3.2. Metod

Traditionellt har logit-modeller ofta använts till nationella färdmedelsvalsmodeller. Eftersom vi introducerar en ny datakälla väljer vi att i denna första artikel att använda oss av beprövade metoder, för att tydligt kunna särskilja effekten av den nya datakällan. Modellstrukturen utgår från logit-modellen i Andersson et al. (2022, Preprint) från Demopan-1 projektet. Där används en nästlad logitmodell för att hantera att mobilnätetsobservationer av resor identifierar både buss- och bilresor som väg i indata från Breyer et al. (2021). Vidare hanteras det faktum att reseärende saknas i mobilnätetsdatan genom att introducera två latenta (dvs dolda) klasser i logit-modellen (tjänsteresa eller privatresa), där indelningen mellan klasserna sker genom skattning av parametrar som kan påverka sannolikheten att tillhöra respektive klass. I de fall där reseobservationen kommer från enkätdata finns redan denna information och färdmedlet sätts direkt till bil, buss, tåg eller flyg, medan reseärende sätts direkt till tjänste- eller privatresa.

Det är känt från tidigare kombinationer av andra datatyper för logit-modeller (Daly and Bradley, 1991) att det behövs en skalparameter i den kombinerade modellen för att ta hänsyn till att variansen i feltermen varierar mellan de två datakällorna. Olika varianser i feltermen i de två datakällorna kan dels bero på att datakällorna har olika typer av mätfel i de oberoende variablerna, men det kan också bero på att det finns olika stora variationer i icke observerade variabler som påverkar valet av färdmedel. Medelvärden av icke observerade variabler som påverkar valet kan också skilja mellan de två datakällorna. Därför använder vi också separata alternativ-specifika konstanter för reseobservationer

från de två datakällorna. Skillnader i sannolikheten att observeras med ett visst färdmedel kan skilja sig mellan enkätdata och mobilnätsdata, vilket är ytterligare ett skäl till att anta olika alternativ-specifika konstanter för datakällorna. Slutligen utvärderar vi en sista modell där nyttofunktionerna skiljer sig mellan mobilnätsobservationer och enkätobservationer, genom att variabler som enbart finns i enkätdata läggs till i enkätdata-nyttfunktionerna men inte i mobilnäts-nyttfunktionerna. Därmed kan vi dra nytta av både den stora mängden mobilnätsdata och de socio-ekonomiska detaljerna från enkätdata.

3.3. Resultat och lärdomar

Det visar sig att de variabler som vi har valt ut för användning endast i enkätdatas nyttofunktioner (ensamresande kvinna, dummy för resesällskap större än 8, körkortsinnehav och antal bilar per körkort i hushållet) inte är korrelerade med resten av modellens variabler, vilket är ett krav för att samskattning ska ge rättvisande resultat. Precis som i den långväga modell som används idag ökar sannolikheten att välja buss för stora resesällskap, sannolikheten att välja tåg eller buss ökar för ensamresande kvinnor och körkortsinnehav och många bilar per körkort i hushållet ökar sannolikheten att välja bil. Att inkludera dessa variabler bidrar till att förbättra modellens precision jämfört med om bara mobilnätsdata hade använts.

De problem med höga restidsvärden för buss och tjänsteresor som identifierades i modellskattningen på enbart mobilnätsdata i Demopan-1-projektet mildras tydligt vid samskattning av de två datakällorna. En bidragande orsak till att restidsvärdena för tjänsteresor minskade var att inkluderandet av enkätdata gjorde det möjligt att få signifikans för latent-klass-parametrar som stämmer bättre med gruppen tjänsteresenärer än då bara mobilnätsdata användes. Detta sker implicit genom att enkätdata förbättrar skattningen av nyttofunktionsparametrarna som bättre stämmer med klasserna tjänste- och privatresor, och på så sätt förbättras skattningen av de latent-klass-parametrar som avgör sannolikheten att tillhöra en viss klass. Restidsvärdena som skattas är dock fortfarande över lag högre än de som används i ASEK (Trafikverket, 2020). Möjliga skäl till skillnaderna är att ASEK-värdena baseras på hypotetiska-val-data (data från hypotetiska val - SP) som normalt sätt ger lägre tidsvärden än faktiska-val-data (avslöjad data - RP) som vi använder här. Det kan både bero på att respondenter inte fullt beaktar sina tidsbegränsningar i SP experiment, men också att kostnaderna för resor i RP data är dåliga, vilket leder till en underskattning av kostnadsparametern, och en överskattning av tidsvärdet (Varela et al., 2018). Tidsvärden från RP-data anses vara mer representativa för verkliga beslutssituationer och borde alltså vara att föredra, men det skulle kräva att skattningsdata för resornas kostnader förbättras. En sådan förbättring skulle också förbättra modellernas prediktioner av effekter av prisförändringar så som förändrade drivmedelspriser, övergång till elbilar och förändrade biljettp priser för kollektivtrafik, tåg och flyg.

Utifrån skattade modeller beräknar vi elasticiteter (egenelasticiteter med avseende på tid och kostnad och korselasticiteter med avseende på tid), varav de flesta är i linje med jämförbara studier. Vi får högre korselasticitet för tjänsteresor med flyg vid förändringar i tågrestid än tidigare och högre tåg- och flyg-korselasticiteter vid förändringar i restid med bil för privatresor än tidigare jämförbara studier.

I dagsläget är tillämpningsområdet för samskattning framför allt långväga resor eftersom noggrannheten i plats inte är särskilt hög för mobilnätsdata. I och med introduktionen av 5G i tätbebyggda områden kan det så småningom även bli en aktuell metod för regionala resor eftersom vissa 5G antenner har kortare räckvidd än GSM, 3G och 4G och därmed har högre geografisk upplösning.

Sammantaget visar vi i artikeln att den föreslagna samskattningsmetoden är giltig, och att det genom samskattning är möjligt att dra nytta av styrkorna hos respektive datakälla.

4. Resultat 2: Att tillämpa maskininlärningsmetoder vid skattning av färdmedelsvalmodell på kombinerade datakällor

Angelica Andersson, Ida Kristoffersson, Maria Börjesson, Clas Rydergren

Manuskript under bearbetning

4.1. Bakgrund och frågeställning

Medan artikeln som beskrevs i avsnitt 1 handlade om hur logit-modeller kan användas för att kombinera enkät- och mobilnätsdata i en färdmedelsvalsmodell, handlar denna artikel om att utreda om maskininlärningsmetoder (ML) kan användas mer effektivt för samma ändamål. Givet att mobilnätsdata skiljer sig från enkätdata (mobilnätsdata saknar indelning mellan bil och buss, saknar reseärende och socio-ekonomisk bakgrund till de resande, och innehåller betydligt fler reseobservationer), så skulle det kunna vara så att det nya dataformatet gör att andra metoder än logit är mer lämpade för uppgiften. Motivationen är densamma, det vill säga att bibehålla hög precision i prognoser för persontransporter även om resvaneundersökningarna har låga svarsfrekvenser.

4.2. Metod

En litteraturgenomgång genomförs för att först bedöma vilken maskininlärningsmetod som har bäst potential som färdmedelsvalsmodell. Därefter beskrivs den kombination av anpassningar av grundmetoden som krävs för att kunna kombinera de två datakällorna, som till viss del skiljer sig åt.

En systematisk litteraturöversikt av ML metoder som färdmedelsvalsmodell genomfördes 2021 av Hillel et al. (2021). Hillel et al. (2021) konstaterar att ML metoder i hög grad har använts som färdmedelsvalsmodeller, och att forskningen ditintills har varit fragmenterad, ibland hamnar i modelleringsmässiga fallgropar, och saknar en standardiserad metod för att jämföra modellers noggrannhet. Över lag verkar ML metoder trots allt prestera bättre i termer av prediktionsnoggrannhet än logit. Även de studier som har publicerats efter Hillel et al. (2021) rapporterar genomgående att ML metoder ger bättre prediktionsnoggrannhet än logit (Ali et al., 2021; D’Cruz et al., 2022; Diallo et al., 2022; García-García et al., 2022; Kong et al., 2022; Li and Kockelman, 2022; Li et al., 2021; Martín-Baos et al., 2021; Mi et al., 2021; Mohd Ali et al., 2022; Pineda-Jaramillo and Arbeláez-Arenas, 2022; Ramanuj et al., 2023; Salas et al., 2022; Truong et al., 2021; Vinayakumar et al., 2023; Zhang et al., 2023). De flesta av studierna som inkluderar Artificiella Neurala Nätverk eller metoder som baseras på ensembler av beslutsträd rapporterar att någon av dessa har bäst noggrannhet. De studier som inkluderar både Artificiella Neurala Nätverk (ANN) och beslutsträd får varierande resultat, där hälften kommer fram till att ANN har bäst noggrannhet (Ali et al., 2021; Salas et al., 2022) medan andra hälften kommer fram till det motsatta (García-García et al., 2022; Tamim Kashifi et al., 2022). Om man utöver metodens noggrannhet också beaktar den beteendemässiga tillförlitligheten i de ekonometriska mått som härleds från modellen visar det sig att ANN är klart bättre lämpad som färdmedelsvalsmodell än beslutsträdbaserade metoder (Martín-Baos et al., 2023). Mycket har dessutom redan gjorts för att ta fram nätverksarkitekturer (motsvarar modellspecifikation inom logit) som lämpar sig för färdmedelsval (Wang et al., 2020). Därför använder vi oss av en ANN i den här studien. Tidigare studier där ML-metoder används som färdmedelsvalsmodeller baseras på enkätdata, men det finns exempel på hur beslutsträdsbaserad-ML har använts för att klassificera reseärende (Burrieza-Galán et al., 2023).

För att hantera att enkätdata har fler oberoende variabler (i form av socio-ekonomi) än mobilnätsdata används binära variabler för att representera att datan saknas. Genom att stegvis förträna (initialskattning) modellen på först bara mobilnätsdata och sedan bara på enkätdata, kan man med hjälp av datautvidgning (data augmentation) av enkätdata dra nytta av att den stora mängden mobilnätsdata kan särskilja mellan flyg, tåg och väg, för att sedan finjustera till att även särskilja bil och buss med hjälp av enkätdata. Dessutom ökar prediktionsförmågan hos modellen i och med att datan kombineras,

eftersom enkätdata dessutom bidrar med socioekonomiska aspekter samtidigt som mobilnätsdatan innehåller många observationer. Vidare används MTLDNN (Multitask learning deep neural networks) för att simultant klassificera reseärende och färdmedelsval. Fördelen med att använda MTLDNN i stället för latent-klass-modellering för att klassificera reseärende är att man inte a priori behöver bedöma vilka variabler som är mest användbara för att särskilja färdmedel respektive reseärende, utan processen är i stället data-driven, och samma variabel kan användas till båda klassificeringsuppgifterna.

4.3. Resultat och lärdomar

ML som alternativ till logit för färdmedelsvalsmodeller är ett starkt växande forskningsfält, och ML visar god potential för ändamålet i termer av noggrannhet. Varianter av ANN visar även god potential i termer av beteendemässig rimlighet, och ANN är därmed den bästa kandidaten bland ML metoderna. Metoder som är standard inom ANN, som förträning och datautvidgning visar sig vara mycket användbara när en del av datamaterialet har högre upplösning (enkätdata innehåller socioekonomiska variabler och särskiljer bil och buss). Ännu en fördel är att man inte behöver anta något a priori om vilka variabler som kommer vara viktigast för att särskilja mobilnätsdatans indelning i reseärende, det kan i stället lösas data-drivet genom MTLDNN.

5. Resultatspridning

Resultaten från projektet Demopan-2 har spridits vid flertalet konferenser och seminarier. Nedan sammanfattas den resultatspridning som skett under projektets gång.

Tabell 1 visar konferenser och seminarier inom Demopan-2.

Tabell 1: Seminarier och konferenser där Demopan-2 har presenterats.

| Datum | Konferens/Seminarium | Presentationens titel | Presentatör(er) |
|-------------------------|---|---|---|
| 2022-06 | Licentiatseminarium vid LiU | Mode choice modelling of long-distance passenger transport based on mobile phone network data | Angelica Andersson |
| 2022-06 | Transportforum 2022 | Skattning av färdmedelsvalsmodell på mobilnätdata | Angelica Andersson |
| 2022-06 | hEART 2022 | Mode choice estimation on mobile network data: challenges and solutions | Angelica Andersson |
| 2022-09 | ETC 2022 | Mode choice logit estimation on mobile network data | Angelica Andersson |
| 2022-10 | Nationell konferens i transportforskning 2022 | Mode choice estimation on mobile network data: challenges and solutions | Angelica Andersson |
| 2023-10 | Nationell konferens i transportforskning 2023 | Mode choice estimation on joint travel survey and mobile phone network data | Angelica Andersson |
| 2023-12 | TRV Webinarium | Resultat från forskningsprojektet Demopan-2 | Angelica Andersson och Ida Kristoffersson |
| 2024-01 (accepterat) | Transportforum 2024 | Long-distance mode choice estimation on joint travel survey and mobile phone network data | Angelica Andersson |

Referenser

- Ali, N.F.M., Sadullah, A.F.M., Majeed, A.P.P.A., Razman, M.A.M., Zakaria, M.A., Nasir, A.F.A., 2021. Travel Mode Choice Modeling: Predictive Efficacy between Machine Learning Models and Discrete Choice Model. *Open Transp. J.* 15, 241–255. <https://doi.org/10.2174/1874447802115010241>
- Andersson, A., 2022. Mode choice modelling of long-distance passenger transport based on mobile phone network data. Linköping University Electronic Press, Linköping Studies in Science and Technology. Licentiate Thesis, ISSN 0280-7971 ; 1936 2022.
- Andersson, A., Börjesson, M., Breyer, N., Daly, A., Engelson, L., Kristoffersson, I., 2022. Mode Choice Latent Class Estimation on Mobile Network Data. Preprint at: <https://doi.org/10.2139/ssrn.4246865>
- Andersson, A., Kristoffersson, I., Daly, A., Börjesson, M., 2023. Long-distance mode choice estimation on joint travel survey and mobile phone network data.
- Breyer, N., Gundlegård, D., Rydergren, C., 2021. Travel mode classification of intercity trips using cellular network data. *Transportation Research Procedia* 52, 211–218.
- Burrieza-Galán, J., Martín, I., Nicolai, F., Monroig, J., Torres, J., Picornell, M., 2023. Identification of long-distance tour purpose through ML-based data fusion techniques for combining mobile network and survey data. *Transportation Research Procedia, XV Conference on Transport Engineering, CIT2023* 71, 93–100. <https://doi.org/10.1016/j.trpro.2023.11.062>
- Daly, A., Bradley, M., 1991. Estimation of Logit Choice Models using Mixed Stated Preference and Revealed Preference Information, in: 6th International Conference on Travel Behaviour. Presented at the 6th International Conference on Travel Behaviour, Québec.
- D’Cruz, J.J.M., Alex, A.P., Manju, V.S., 2022. MODE CHOICE ANALYSIS OF SCHOOL TRIPS USING RANDOM FOREST TECHNIQUE. *Arch. Transp.* 63, 39–48. <https://doi.org/10.5604/01.3001.0015.9175>
- Diallo, A.O., Lozenguez, G., Doniec, A., Mandiau, R., 2022. Estimation of minority modes of transportation based on machine learning approach, in: *Procedia Comput. Sci.* Presented at the *Procedia Computer Science, Elsevier B.V.*, pp. 265–272. <https://doi.org/10.1016/j.procs.2022.03.036>
- García-García, J.C., García-Ródenas, R., López-Gómez, J.A., Martín-Baos, J.Á., 2022. A comparative study of machine learning, deep neural networks and random utility maximization models for travel mode choice modelling, in: Coelho M. (Ed.), *Transp. Res. Procedia*. Presented at the *Transportation Research Procedia, Elsevier B.V.*, pp. 374–382. <https://doi.org/10.1016/j.trpro.2022.02.047>
- Hillel, T., Bierlaire, M., Elshafie, M.Z.E.B., Jin, Y., 2021. A systematic review of machine learning classification methodologies for modelling passenger mode choice. *Journal of Choice Modelling* 38, 100221. <https://doi.org/10.1016/j.jocm.2020.100221>
- Kong, M.S., Lee, J.-H., Park, B.H., 2022. Establishment of Travel Mode Choice Model in Intercity Passenger Travel Using Deep Neural Networks. *J. Korean Soc. Railw.* 25, 354–366. <https://doi.org/10.7782/JKSR.2022.25.5.354>
- Li, W., Kockelman, K.M., 2022. How does machine learning compare to conventional econometrics for transport data sets? A test of ML versus MLE. *Growth Change* 53, 342–376. <https://doi.org/10.1111/grow.12587>
- Li, X., Wang, Y., Wu, Y., Chen, J., Zhou, J., 2021. Modeling Intercity Travel Mode Choice with Data Balance Changes: A Comparative Analysis of Bayesian Logit Model and Artificial Neural Networks. *J Adv Transp* 2021. <https://doi.org/10.1155/2021/9219176>

- Martín-Baos, J.Á., García-Ródenas, R., Rodríguez-Benitez, L., 2021. Revisiting kernel logistic regression under the random utility models perspective. An interpretable machine-learning approach. *Transp. Lett.* 13, 151–162. <https://doi.org/10.1080/19427867.2020.1861504>
- Martín-Baos, J.Á., López-Gómez, J.A., Rodríguez-Benitez, L., Hillel, T., García-Ródenas, R., 2023. A prediction and behavioural analysis of machine learning methods for modelling travel mode choice. *Transp. Res. Part C Emerg. Technol.* 156. <https://doi.org/10.1016/j.trc.2023.104318>
- Mi, X., Wang, S., Shao, C., Zhang, P., Chen, M., 2021. Resident travel mode prediction model in Beijing metropolitan area. *PLoS ONE* 16. <https://doi.org/10.1371/journal.pone.0259793>
- Mohd Ali, N.F., Mohd Sadullah, A.F., Majeed, A.P.P.A., Mohd Razman, M.A., Choong, C.S., Muazu Musa, R., 2022. Forecasting Daily Travel Mode Choice of Kuantan Travellers by Means of Machine Learning Models, in: Ab. Nasir A.F., Ibrahim A.N., Ishak I., Mat Yahya N., Zakaria M.A., P. P. Abdul Majeed A. (Eds.), *Lect. Notes Electr. Eng. Presented at the Lecture Notes in Electrical Engineering*, Springer Science and Business Media Deutschland GmbH, pp. 979–987. https://doi.org/10.1007/978-981-33-4597-3_89
- Pineda-Jaramillo, J., Arbeláez-Arenas, Ó., 2022. Assessing the Performance of Gradient-Boosting Models for Predicting the Travel Mode Choice Using Household Survey Data. *J. Urban Plann. Dev.* 148. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000830](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000830)
- Ramanuj, P., Varia, H., Shah, A., Jain, A.M., 2023. DEVELOPMENT OF MODE CHOICE BEHAVIOR MODEL FOR INTER-REGIONAL PUBLIC TRANSPORT– A CASE STUDY OF INDIA. *Suranaree J. Sci. Technol.* 30.
- Salas, P., De la Fuente, R., Astroza, S., Carrasco, J.A., 2022. A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. *Expert Sys Appl* 193. <https://doi.org/10.1016/j.eswa.2021.116253>
- Stopher, P.R., Greaves, S.P., 2007. Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice, Bridging Research and Practice: A Synthesis of Best Practices in Travel Demand Modeling* 41, 367–381. <https://doi.org/10.1016/j.tra.2006.09.005>
- Tamim Kashifi, M., Jamal, A., Samim Kashafi, M., Almoshaogeh, M., Masiur Rahman, S., 2022. Predicting the travel mode choice with interpretable machine learning techniques: A comparative study. *Travel Behav. Soc.* 29, 279–296. <https://doi.org/10.1016/j.tbs.2022.07.003>
- Trafikverket, 2020. *Analysmetod och samhällsekonomiska kalkylvärden för transportsektorn: ASEK 7.0.*
- Truong, T.M.T., Ly, H.-B., Lee, D., Pham, B.T., Derrible, S., 2021. Analyzing travel behavior in Hanoi using Support Vector Machine. *Transp. Plann. Technol.* 44, 843–859. <https://doi.org/10.1080/03081060.2021.1992178>
- Varela, J.M.L., Börjesson, M., Daly, A., 2018. Quantifying errors in travel time and cost by latent variables. *Transportation Research Part B: Methodological* 117, 520–541. <https://doi.org/10.1016/j.trb.2018.09.010>
- Vinayakumar, G., Alex, A.P., Manju, V.S., 2023. A Comparison of KNN Algorithm and MNL Model for Mode Choice Modelling. *Eur. Transp. Trasporti Eur.* <https://doi.org/10.48295/ET.2023.92.3>
- Wang, S., Mo, B., Zhao, J., 2020. Deep neural networks for choice analysis: Architecture design with alternative-specific utility functions. *Transportation Research Part C: Emerging Technologies* 112, 234–251. <https://doi.org/10.1016/j.trc.2020.01.012>
- Zhang, H., Zhang, L., Liu, Y., Zhang, L., 2023. Understanding Travel Mode Choice Behavior: Influencing Factors Analysis and Prediction with Machine Learning Method. *Sustainability* 15. <https://doi.org/10.3390/su151411414>

VTI är ett oberoende och internationellt framstående forskningsinstitut inom transportsektorn. Vi bedriver forskning och utveckling för att förbättra kunskapen om infrastruktur, trafik och transporter. Genom vårt arbete bidrar vi till att nå Sveriges transportpolitiska mål för tillgänglighet, säkerhet, miljö och hälsa.

Vi utför forskning på uppdrag inom alla transportslag och arbetar i en tvärvetenskaplig organisation. Den kunskap vi genererar ger viktig information till aktörer inom transportsektorn och används ofta direkt i nationell och internationell transportpolitik.

Utöver forskning erbjuder vi utredningar, rådgivning samt olika mät- och provningstjänster. På VTI har vi avancerad forskningsutrustning av olika slag och världsledande körsimulatorer. Vi har även ackrediterade laboratorier för vägmaterial och krocksäkerhetstestning.

Biblioteket vid VTI är en nationell resurs som samlar in och sprider information om svensk transportforskning. Utöver frågeservice och lån erbjuds tjänster såsom informationssökning, omvärldsbevakning och strukturering av publikationer och projekt på en webbplats.

I Sverige samarbetar VTI med universitet och högskolor som bedriver relaterad forskning och utbildning. Vi deltar regelbundet i internationella forskningsprojekt, främst i Europa, och är aktiva inom internationella nätverk och allianser. Vi är cirka 240 medarbetare och finns i Linköping, Stockholm, Göteborg och Lund.

vti

Statens väg och transportforskningsinstitut · www.vti.se · +46 (0)13-20 40 00
