

LOPPURAPORTTI

Tiedolla johtaminen -2023

Vesiliikenteen kehittäminen ja
kysyntäjoustoihin vastaaminen
tietopohjaisesti

Helsinki



Tukea digitalisaatiokokeiluihin
kaupungin työntekijöille

Vesiliikenteen kehittäminen ja kysyntäjoustoihin vastaaminen tietopohjaisesti on vasta alussa. Kokonaisvaltainen kehitys voisi luoda kestävämpää ja responsiivisempaa vesiliikennettä tulevaisuudessa.

**Tapio Rossi,
Helsingin kaupunki**

**Aleksi Könönen, Pauline Ranta, Sami Räsänen
Futurice**

Sisällysluettelo

1. Tiivistelmä
2. Kokeilun tavoitteet
3. Kokeilun keskeiset opit
4. Kokeilun eteneminen
5. Kokeilun tuotokset
6. Opit kokeiltavan ratkaisun tai toimintatavan mahdollisuuksista
7. Opit asiakkaiden tai palvelun käyttäjien tarpeista
8. Opit ratkaisun kehittämisestä teknisesti
9. Opit kokeilemisestä yleensä
10. Opit kokeiluprojektin arjen pyörittämisestä
11. Kokeilun tekninen ympäristö
12. Kokeilun data
13. Jatkopäätökset ja -ideat

1. Tiivistelmä

● Kokeilun tavoitteet

- **Kokeilun tarkoituksena on selvittää data-analyysin avulla, mitkä tekijät vaikuttavat vesiliikenteen matkustajamääriin.** Suurin huomio kiinnitettiin sääolosuhteiden vaikutukseen. Kokeilun edetessä käsiteltiin myös mahdollisuutta ottaa käyttöön joustava aikataulutus ja hinnoittelu.

● Kokeilun opit

- **Helsingin vesiliikenteen kävijämääriin vaikuttaa merkittävästi sää.** Vaikka näitä kävijämääriä pystytään ennustamaan tietyn menetelmän avulla, ennustetarkkuudessa on vielä parantamisen varaa. On olennaista ymmärtää, että sää ei yksinään määrää kävijämääriä, vaan toimii yhtenä vaikuttavana tekijänä. **Suurin haaste tarkkuudessa liittyy datan yksinkertaisuuteen ja sen epätarkkuuteen.**

● Suositukset jatkotoimenpiteiksi

- **Datan keräyksen kehittäminen liikennöitsijöillä.** Tämä voi sisältää esimerkiksi uusien teknologisten ratkaisujen käyttöönottoa, nykyisten menetelmien tehostamista tai käytäntöjen muuttamista niin, että saadaan laadukkaampaa, tarkempaa ja hyödyllisempää tietoa liikennöintitoiminnasta
- **Yhteistyön laajentaminen ja kehittäminen:** laajennetaan yhteistyötä viranomaisten ja liikennöitsijöiden välillä ja kehitetään molempia hyödyttäviä kilpailutuksia
- **Teknologisten ratkaisujen päivittäminen:** otetaan käyttöön ratkaisuja, jotka mahdollistavat paremman datan keräämisen ja analysoinnin vesiliikenteen kontekstissa
- **Käyttäjäkokemuksen parantaminen:** keskitetään ponnistelut datankeruun lisäksi myös käyttäjäkokemuksen parantamiseen esimerkiksi liikenteen seuranta- ja ilmoituspalveluiden avulla.

2. Kokeilun tavoitteet

- Ongelman kuvaus
 - Helsingin vesiliikenteessä on havaittu tarvetta kehittää mm. ympäristöystävällisyyttä, asiakaskokemusta ja kustannustehokkuutta. Tällä hetkellä vesiliikenteen toiminnasta saatava tieto ei kuitenkaan riitä tietopohjaiseen kehitysohjon ja tiedolla johtamiseen.
- Rajaus
 - Tutkitaan lämpötilan ja sademäärän vaikutusta kävijämääriin vuonna 2023. Saatujen oppien perusteella voidaan tulevaisuudessa laajentaa datan hyödyntämistä myös muissa osa-alueissa
- Oletukset
 - Kokeilun hypoteesina on, että säällä on vaikutus matkustajamääriin.
- Tavoitteet
 - Selvittää, millainen vaikutus säällä on kävijämääriin.

3. Kokeilun keskeiset opit

- Tapahtuiko kokeilussa niin kuin olit oletanut?
 - Kokeilun alkuperäinen hypoteesi osoittautui todeksi, ja tämä sai meidät syventymään aiheeseen tarkemmin. Kokeilun aikana kuitenkin huomattiin tarve laajemmalle kehitykselle, erityisesti liittyen data-alustan ja liikennöitsijäsopimusten päivitykseen.
- Miten kokeilu onnistui suhteessa asetettuihin tavoitteisiin?
Yhteenveto kokeilun toteutumisesta tavoitteisiin peilaten.
Yksityiskohtaiset tulokset esitellään seuraavilla dioilla.

Kokeilu sujui menestyksekkäästi suhteessa asetettuihin tavoitteisiin. Seuraavassa on yhteenveto kokeilun toteutumisesta tavoitteisiin peilaten:

- Kokeilu vahvisti alkuperäisen hypoteesimme ja osoitti, että se oli validi.
- Kokeilun myötä saimme syvällisempää ymmärrystä vesiliikenteen kehittämisen mahdollisuuksista ja liikennöitsijöiden näkökulmista tietopohjaiseen kehittämiseen käytännössä.
- Vaikka kokeilu oli onnistunut, havaittiin samalla tarve laajemmille kehitysohjelmille.

4. Kokeilun eteneminen

Projektin alkuvaiheessa keräsimme vesireiteiltä monipuolista dataa suoraan liikennöitsijöiltä. Saadut tiedot toimitettiin Excel-muodossa, jossa oli erilaisia formaatteja. Säädata puolestaan hankittiin Ilmatieteen laitoksen tarjoamasta käyttöliittymästä. Yhdistimme nämä datasetit Pythonilla ja teimme alustavan analyysin sään ja kävijämäärien välisestä suhteesta.

Tulosten perusteella vahvistui havainto, että säällä on merkittävä vaikutus kävijämääriin. Kokeilussa ensin testasimme regressiomallin sovittamista dataan, mutta tulokset eivät olleet tyydyttäviä. Epäilimme, että tämä johtui regressiomalleista, jotka eivät pystyneet tunnistamaan tiettyjä trendejä, kuten kasvavaa liikennettä viikonloppuisin.

Tämän jälkeen siirryimme luomaan ennustavaa aikasarjamallia. Aikasarjamallin valmistuttua palasimme takaisin regressiomalleihin, joita tutkittiin tarkemmin. Testasimme matkustajamäärien ennustamista viidellä eri regressiomallilla ja analysoimme tuloksia.

Laajemmassa analyysissä vertailimme erilaisten mallien suorituskykyä kävijämäärien ennustamisessa. Tämän jälkeen järjestimme työpajan liikennöitsijöiden kanssa, jonka tavoitteena oli kartoittaa jatkokehitysmahdollisuuksia.

Lopuksi koostimme raportin tuloksista, jossa esiteltiin kattavasti projektin vaiheet, havainnot ja suositukset jatkotoimenpiteiksi.

5. Kokeilun tuotokset

- Kokeilussa on käytetty kahdenlaisia ennustavia malleja, regressiomalleja ja yhtä aikasarjamallia (SARIMAX).

Regressioanalyysi on tilastollinen menetelmä, jota käytetään yhden muuttujan (kuten matkustajien) arvon ennustamiseen toisen tai useamman muuttujan (kuten sään) perusteella. Se auttaa ymmärtämään, kuinka suuresti riippuvat muuttujat vaikuttavat ennustettavaan muuttujaan ja onko niiden välinen suhde merkittävä.

SARIMAX, lyhenne sanoista "Seasonal AutoRegressive Integrated Moving Average with eXogenous variables", on aikasarja-analyysin menetelmä, joka ottaa huomioon sekä aikasarjan itsensä että ulkopuoliset muuttujat ennustaessaan tulevia arvoja. Tämä malli on erityisen hyödyllinen aikasarjojen analysoinnissa, joissa esiintyy kausivaihtelua ja muita trendejä.

Näiden mallien pääero ilmenee niiden soveltamisalueessa: regressioanalyysi on yleisemmin käytetty yleistyökalu, joka soveltuu monenlaisiin tilanteisiin. Toisaalta SARIMAX on erikoistuneempi ja sopii parhaiten aikasarja-analyysiin, jossa otetaan huomioon sekä sisäiset että ulkoiset tekijät. Regressioanalyysin vahvuus piilee sen yksinkertaisuudessa ja monipuolisuudessa. Se ei kuitenkaan välttämättä sovellu monimutkaisiin aikasarjadatan ennustamisiin, joissa kausivaihtelut ja muut aikasarjan ominaisuudet ovat merkittäviä. Toisaalta SARIMAX tarjoaa tarkempia ennusteita tällaisissa tapauksissa, mutta se on monimutkaisempi ja vaatii tarkempaa tietoa aikasarjan luonteesta.

5. Kokeilun tuotokset

Mitä tulee tietää kun tuotoksia analysoidaan?

- R^2 on korrelaatiokerroin, joka kuvaa mallin (säätietojen) kykyä selittää muuttujan (matkustajamäärien) vaihtelua. Esimerkiksi, kun R^2 -arvo on 0.2, noin 20% matkustajamäärien vaihtelusta voidaan selittää säätiedoilla. Negatiivinen R^2 -arvo osoittaa, että valittu malli sopii huonommin dataan kuin yksinkertainen malli, joka ennustaa vain riippuvan muuttujan keskiarvoa. Tämä viittaa siihen, että malli ei selitä havaittua vaihtelua lainkaan tai tekee sen erittäin heikosti.
- RMSE, eli keskivirhe tai virhetermi, edustaa käytännössä ennustevirhettä. Se kuvaa, kuinka paljon ennustettu matkustajamäärä keskimäärin poikkeaa toteutuneesta. Ylisovittaminen tapahtuu, kun tilastollinen malli oppii koulutusdatan satunnaisvaihtelut ja kohinan niin tarkasti, että se menettää kykynsä ennustaa uusia, näkemättömiä tietoja tarkasti.
- Datapiste on yksittäinen tiedon ilmentymä aikasarjadataassa. Tässä kokeilussa, kun saatava data on päiväkohtaista, jokainen päivä tuottaa yhden datapisteen.

5. Kokeilun tuotokset

Lyhyt kuvaus käytetyistä regressiomalleista:

Ridge-regressio:

- Käytetään mallin monimutkaisuuden rajoittamiseen ja ylisovittamisen välttämiseen.
- Lisää virhetermiin pienen kertaluvun, joka rangaistuksena vähentää suuria painoarvoja.

Lasso-regressio:

- Valitsee tärkeät muuttujat ja eliminioi epäolennaiset kertoimet mallista.
- Asettaa vähämerkityksellisten muuttujien kertoimet nolaksi, mikä yksinkertaistaa mallia ja ehkäisee ylisovittamista.

Decision Tree -regressio:

- Luo päätöspuun jakamalla datan pienempiin osiin ominaisuuksien arvojen perusteella.
- Pystyy tekemään ennusteita muuttujasta päätöspuun avulla.

Random Forest -regressio:

- Parantaa ennustuskykyä yhdistämällä useita päätöspuita, jotka on koulutettu eri datan alijoukoilla.
- Vähentää mallin varianssia ja tarjoaa parempaa suorituskykyä.

Support Vector -regressio:

- Etsii optimaalisen ennustavan funktion luomalla ja optimoimalla rajapinnan, joka maksimoi marginaalin lähimmän datapisteen ja ennusteen välillä.

5. Kokeilun tuotokset

Ennustavien mallien vertailu

	Kaikki		Liuskasaari		Harakka		Pihlajasaari		Uunisaari		Merisatama	
	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
Ridge Regression	0.292	132.33	0.1971	9.946356117	-0.0462	32.91640928	0.376	196.1960244	-0.8042	5.368426213	0.347	32.49
Lasso Regression	0.302	131.35	0.1855	10.01848292	-0.0675	33.24906014	0.359	198.8581152	-0.007	4.009987531	0.3244	33.04
Decision Tree Regressor	-0.545	195.44	-0.9926	15.66971602	-1.0508	46.08535559	0.3706	197.0459084	-1.6295	6.480740698	0.5802	26.05
Random Forest Regressor	-0.911	217.41	-0.3181	12.74480286	-0.3862	37.88984033	0.3064	206.8528221	-0.6151	5.07937004	0.4128	30.81
Support Vector Regression	-0.013	158.28	-0.0959	11.62067124	0.0208	31.84572185	-0.3365	287.1293785	-0.0602	4.114608122	0.0052	40.10
SARIMAX	-23.01	38.21										

R² = Selityskerroin, kuvaa sitä, kuinka hyvin mallissa tekijät (säätiedot) kuvaavat muuttujan (matkustajamäärien) vaihtelua. Esimerkiksi kun R² arvo on 0.2, suurinpiirtein 20% matkustajamäärien vaihtelusta pystytään selittämään säätiedoilla. Negatiivinen R²-arvo tarkoittaa, että valittu malli sopii dataan huonommin kuin yksinkertainen malli, joka ennustaa vain riippuvan muuttujan keskiarvoa, viitaten siihen, että malli ei selitä havaittua vaihtelua ollenkaan tai tekee sen erittäin huonosti.

RMSE = Keskiarvo käytännössä virhemarginaali. Kuinka paljon ennustettu matkustajamäärä keskimäärin eroaa toteutuneesta. HUOM! Vaikka esimerkiksi Harakassa RMSE arvot ovat pienempiä kuin esimerkiksi Pihlajasaarella, se ei tarkoita, että Pihlajasaaren malli olisi huonompi, sillä Pihlajasaaren kävijämäärät ovat myös suuremmat.

5. Kokeilun tuotokset

- Mallien testeistä havaittiin, että saatavilla olevalla datasetillä aikasarja-analyysimalli ei suoriutunut hyvin. Datasetti oli melko pieni (659 datapistettä) ja käsitti vain yhden laajemman trendikauden (vain yhden vuoden tiedot). Aikasarjaennustemalli olisi voinut menestyä paremmin tarkemmalla ja laajemmalla datamäärällä, mikä olisi mahdollistanut päivä-, viikko- ja kuukausitrendisykliä tarkan oppimisen.
- SARIMAX-malli vaatii usein hienosäätöä tarkkojen ennusteiden tekemiseksi, mutta tämä ei olisi ollut hyödyllistä, sillä saatavilla olevilla datamäärillä ei olisi voitu luoda luotettavaa ennustemallia.
- Eri regressiomallien välillä esiintyi jonkin verran vaihtelua eri reittien suhteen johtuen asiakaskäyttäytymisen erilaisuuksista. Esimerkiksi Pihlajasaaren kävijämääriä voitiin ennustaa parhaiten useimmilla käytetyillä malleilla, kun taas Uunisaaren kävijämäärään sääolosuhteet eivät näyttäneet vaikuttavan. Uunisaaren datapisteiden määrä oli myös pienin (88 datapistettä), mikä vaikutti mallien heikkoon tarkkuuteen.
- Ennustavat mallit toimivat parhaiten reiteillä, joilla on suurempia kävijämääriä, koska näillä reiteillä mallien on helpompi oppia ja ennustaa, kun datan suhteellinen vaihtelu ei ole niin suurta.
- Yleisesti ottaen Lasso-regressiomalli suoriutui parhaiten kaikista kokeilluista malleista, mutta sen keskivirhe oli yleisesti korkea. On tärkeää huomata, että datan laadun parantuessa myös sopivin ennustemalli voi muuttua. Kerätessä pidemmältä aikaväliltä ja tarkemmalla tasolla dataa huolellisesti hienosäädetty SARIMAX-malli todennäköisesti päihittää regressiomallit.

5. Kokeilun tuotokset

Date

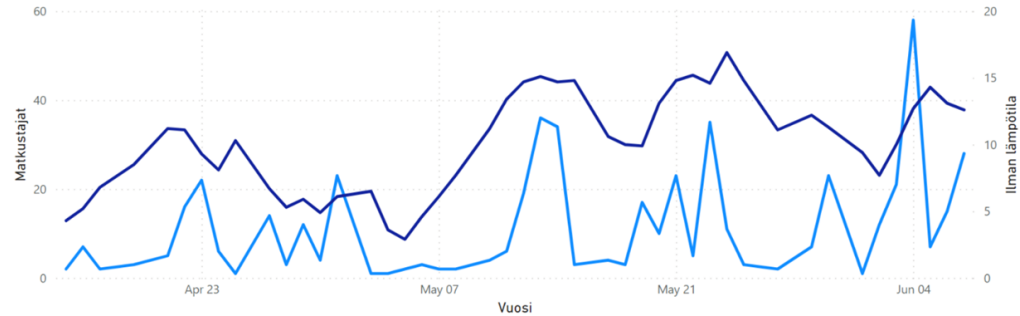
3/22/2023 6/7/2023

Route

- Harakka
- Liuskasaari
- Merisatama
- Pihlajasaari
- Uunisaari

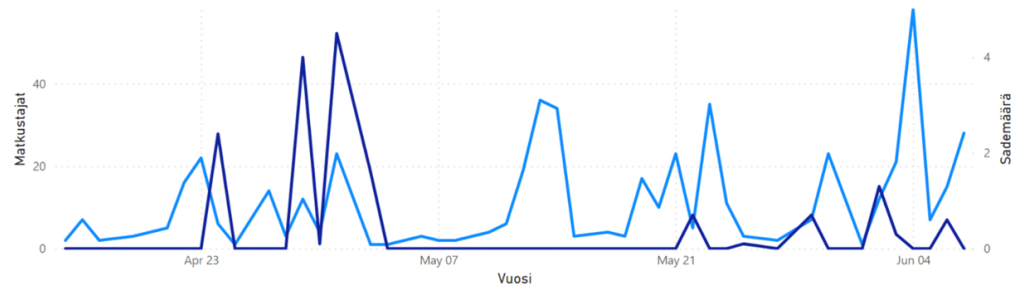
Matkustajat ja Ilman lämpötila

● Matkustajat ● Ilman lämpötila



Matkustajat ja Sademäärä

● Matkustajat ● Sademäärä



6. Opit kokeiltavan ratkaisun tai toimintatavan mahdollisuuksista

Kokeilun aikana saatiin uusia näkökulmia ongelmaan. Tiedostettiin, että ongelma ei ole ainoastaan tietotekninen, vaan siihen liittyy myös sopimusteknisiä komponentteja. Tämä oivallus auttoi laajentamaan ongelman ymmärrystä ja keskittymään ratkaisujen etsimiseen monipuolisemmin.

Jotta alkuperäinen ongelma voitaisiin ratkaista skaalautuvasti ja tuotannollistaa, kerätyn datan tulisi olla tarkkaa tietoa matkakohtaisista matkustajamääristä. Lisäksi tämän datan tulisi olla saatavilla yhteisellä data-alustalla.

Kokeilusta opittiin, että mitä pidempään tällaista dataa kerätään, sitä tarkempia, parempia ja vakaampia ennustemalleja voidaan luoda. Tämä korostaa tarvetta systemaattiselle ja jatkuvasti kehittyvälle datankeruulle osana ratkaisun skaalaamista.

7. Opit asiakkaiden tai palvelun käyttäjien tarpeista

Toimeksiannossa liikennöitsijöiden osallistuminen oli tärkeä osa prosessia. Liikennöitsijät ovat niitä, jotka tarjoavat liikennepalveluja ja keräävät dataa matkustajamääristä. Heidän roolinsa tässä kontekstissa liittyi datan toimittamiseen analyysiä varten. Liikennöitsijät olivat myös avainasemassa tarjoamalla tietoa muista relevantista tiedosta, joka vaikuttaa liikennepalveluiden käyttöön.

Liikennöitsijäpalveluiden tilaaja eli Helsingin kaupunki toimii tiedon käyttäjänä toisella tasolla. Helsingin kaupunki, kuten matkustajatkin, hyötyy suoraan palveluiden laadusta, luotettavuudesta ja tehokkuudesta. Keskeisin oppi tilaajaan toimintaan liittyen on, että hankinnan kohdetta, kilpailutuksia ja liikennöitsijäsopimuksia kehittämällä on mahdollista luoda elinvoimaista merellistä Helsinkiä sekä kannattavaa liiketoimintaa palveluita tarjoaville yrityksille ja sitä kautta parempia loppukäyttäjäpalveluita asukkaille ja vierailijoille.

Tulevaisuuden kuluttajat odottavat luotettavia ja ajantasaisia tietoja liikennöitsijäpalveluiden saatavuudesta, aikatauluista ja mahdollisista muutoksista. Käyttäjien muuttuviin tarpeisiin voidaan vastata tässä raportissa kuvatuilla keinoilla.

7. Opit asiakkaiden tai palvelun käyttäjien tarpeista

Järjestimme liikennöitsijöiden etätyöpajan 15.11.2023, johon kutsuttiin kymmenen liikennöitsijää: Aava Lines, Juva Shipping, JT-line, Eurowork, FRS-Finland, Royal Line, Strömma, Sunlines, Ihalines ja Reposalmen lautturit.

Työpajaan osallistuivat:

Stömma Suomi Oy

JT-Line Oy

Royal Line Oy

Aava Lines Oy

Reposalmen lautturit Oy

Työpajan tavoitteena oli saada vesiliikenteen liikennöitsijöiden palautetta tulevaisuuden suuntaan, jossa kysyntäjoustoja voitaisiin mitata ja ennustaa paremmin kaikkia osapuolia hyödyttävällä tavalla. Futurice fasilitoi keskustelun seuraavan agendan mukaisesti:

- Työpajan avaus: tavoite ja juoksutuksen läpikäynti
- Data-analyysin esittely
- Keskustelukierros: mitkä tekijät vaikuttavat matkustajamääriin, jo tunnistettujen lisäksi?
- Ratkaisut: miten datankeruuta kannattaisi kehittää?
- Yhteenveto

7. Opit asiakkaiden tai palvelun käyttäjien tarpeista

Keskeiset työpajasta nousseet teemat:

Tietojenkeruun kehittäminen

- Keskustelu keskittyi tietojenkeruumenetelmien parantamiseen paremman päätöksenteon mahdollistamiseksi

Tulevat investoinnit

- Kysymys siitä, kuka tekee tulevat investoinnit, tilaaja vai liikennöitsijä, nousi esille

Yhteistyön esteet

- Keskusteltiin yhteistyön haasteista, mukaan lukien sopimusjärjestelyt ja käytännön toteutus

Matkustajamääriin vaikuttavat tekijät

- Keskusteltiin erilaisista tekijöistä, jotka vaikuttavat matkustajamääriin, kuten sää, vuodenaika, tapahtumat ja juhlapyhät

LATI -järjestelmä

- LATI Web on HSL:n tilaaman lauttaliikenteen tilastointisovellus. Järjestelmään kirjataan matkustajamäärät, kuljetetut ajoneuvot sekä matkustajat ja / tai ajoneuvot, jotka eivät mahtuneet kyseiselle lähdölle.
- Lati-järjestelmä mainittiin hyödyllisenä työkaluna tilaajalle, mutta ei välttämättä liikennöitsijälle.

Kapasiteetin Allokaatio

- Korostettiin tarvetta tehokkaalle alustalle kapasiteetin allokoimiseksi ja kapasiteetin irrottamisen haasteista.

7. Opit asiakkaiden tai palvelun käyttäjien tarpeista

Yhteenvetona Liikennöitsijöiden kommentit osoittavat tarvetta parempaan investointien hallintaan, kysyntäperusteiseen liikennöintimalliin ja mahdollisuuteen ennustaa ja reagoida matkustajamääriin.

Sopimusasiat ja kapasiteetin irrottaminen ovat tällä hetkellä keskeisiä haasteita. Tarvetta on myös paremmalle alustalle ja koordinoinnille eri osapuolten välillä.

Keskeiset työpajasta nousseet jatkokehitysaihiot:

- mikäli liikennöintimallia muutetaan kysyntäperusteiseksi, tulisi se tehdä liiketoiminnallisesti kannattavalla sopimusmallilla
- sopimukselliset haasteet: kilpailutukset sekä niiden vaatimukset/kriteerit ja vertailuperusteet, sopimusten pituus ja muiden liikennöitsijöiden hallinnan ulkopuolella olevien tekijöiden vaikutus yhteistyöhön on merkittävä
- keskustelua kapasiteetin irrottamisen haasteista ja hyvän alustan tarpeesta kannattaa jatkaa
- Lati-järjestelmän käytön mahdollinen laajennus on tutkimisen arvoinen vaihtoehto

8. Opit ratkaisun kehittämisestä teknisesti

Analyttiset mallit ja ennustemallit ovat parhaimmillaan yhtä hyviä kuin niiden analysoima data. Toisin sanoen, ennustemallit ovat riippuvaisia saatavilla olevasta datasta. Laadukas ja monipuolinen data mahdollistaa mallin paremman oppimisen ja antaa sille vahvemman perustan ennustamiseen.

Ennustemallien suorituskyky riippuu myös siitä, kuinka ajantasaista dataa on käytettävissä. Vanhentunut tai epätarkka data voi johtaa virheellisiin ennusteisiin. Jatkuvasti päivitetty ja oikea-aikainen data parantaa mallin luotettavuutta ja tarkkuutta.

Mallit tarvitsevat kattavan datan, joka heijastaa laajasti niitä tekijöitä, joita ne pyrkivät ennustamaan. Puutteellinen tai rajoitettu data voi johtaa mallin puutteellisiin tuloksiin tai jopa virheellisiin päätelmiin.

Monimuotoinen data, joka kattaa useita näkökulmia ja liittyy eri osa-alueisiin, auttaa mallia ymmärtämään ilmiötä monipuolisemmin. Monitieteellinen data voi paljastaa monimutkaisempia suhteita, joita yksipuolinen data ei välttämättä näytä.

Mallien jatkuvaa oppimista voidaan parantaa tuomalla niihin jatkuvasti uutta ja relevanttia dataa. Tämä on erityisen tärkeää tilanteissa, joissa olosuhteet tai käyttäytymismallit muuttuvat ajan myötä.

9. Opit kokeilemisesta yleensä

- Mikä toimi hyvin, mitä jatkossa kannattaa tehdä samalla tavalla?

Projektissa oli selkeästi määritelty tavoitealue, joka jopa ylitettiin odotusten. Erityisen positiivisesti korostui projektin alkuvaiheessa säännölliset viikottaiset tapaamiset asiakkaan kanssa. Näissä tapaamisissa käytiin läpi saavutettua edistystä ja suunniteltiin tulevia toimenpiteitä. Tämä käytäntö varmisti, että projekti eteni oikeaan suuntaan ja vältettiin tarpeetonta työtä.

- Mikä ei toiminut odotetusti? Mitä jatkossa kannattaa muuttaa tai tehdä toisin?

Projektin alkuvaiheessa olisi hyödyllistä syventyä vesiliikenteen dataekosysteemin yksityiskohtiin ja suunnata kokeilu kohti konkreettista ratkaisua, joka tuottaisi välittömän askeleen eteenpäin.

- Mitä opit muilta Kokeilukiihdyttämössä mukana olleilta tiimeiltä, kaupungin asiantuntijoilta tai valmentajilta?

Eri tilaisuuksissa on ollut runsaasti mahdollisuuksia jakaa tietoa, ideoita ja kokemuksia. Tämä on rikastuttanut tietämyspohjaamme ja mahdollistaa parhaillaan monipuolisemman näkökulman asioihin.

10. Opit kokeiluprojektin arjen pyörittämisestä

- Kuinka paljon kokeilu vei lopulta aikaa ja miten ajankäyttö jakaantui kokeilun aikajanalla?

Tiimitämme on kulunut 15.12.2023 mennessä projektiin 94h (n. 12,5 htp). Eniten aikaa meni datan käyttöönottoon, aikasarja-analyysiin, Kokeilukiihdyttämön työpajoihin, välikatselmoiteihin ja näiden esitehtäviin.

- Miten hyvin kokeilun budjetti riitti?

Toteutunut tuntihinta on alempi kuin “normaalissa” toimeksiannossa. Toisaalta tähän vaikutti Futuricen päätös osallistaa liikennöitsijöitä suunniteltua laajemmin prosessiin.

- Mitä resursoinnista opittiin?

Projektin suunnittelussa olisi olennaista aloittaa asiakasymmärrysvaiheella tai hypoteesin validoinnilla sekä kokeilun liiketoiminnallisella validoinnilla. Resurssien ollessa rajalliset, on tärkeää tehdä tämä vaihe riittävällä tasolla. Kokeilukiihdyttämön prosessin voisi harkita kaksivaiheiseksi, sisältäen asiakas/käyttäjävalidoinnin ja päätöspisteen ennen siirtymistä kehitysvaiheeseen. Osallistuminen välikatselmoihiin tulisi harkita huolellisesti, sillä niiden aikavaatimus saattaa olla liiallinen suhteessa saavutettuun hyötyyn.

11. Kokeilun tekninen ympäristö

- Millainen oli kokeilun toteutusympäristö? Kuvaa tekniset ratkaisut kuten esim. pilvipalvelualusta, tietokannat, sovellukset, algoritmit

Tekninen toteutus tehtiin Jupyter Notebooksilla käyttäen Pythonia ja useita Pythonkirjastoja.
Datan yleiskatsauksessa käytettiin myös kevyttä PowerBI -raporttia

- Mitä teknisessä ympäristössä olisi hyvä ottaa huomioon, jos kokeilu tehtäisiin nyt uudelleen?

Tekninen puoli soveltui erinomaisesti lyhyeen data-analyysiin, jonka tulos on pelkkä raportti eikä vaadi tuotannollistettavaa tuotetta.

12. Kokeilun data

- Mitä kokeilun datasta ja sen käsittelystä opittiin?

Ennustusta on mahdollista tehdä, mutta datassa on tällä hetkellä ongelmia sen laajuuden ja tarkkuuden suhteen.

Datan puutteet vaikuttavat suoraan ennusteiden tarkkuuksiin. Matkustajamäärien ennustamisen merkittävin havainto on, että vaikka matkustajamääriä pystytään ennustamaan erittäin tarkasti selityskertoimilla, käytettävissä olevan datan määrällä on merkittävä vaikutus virhemarginaaliin. Keskimäärin noin 30% matkustajamäärien vaihtelusta voidaan selittää säätiedoilla.

Datan käsittelyyn liittyvät opit olivat, että nykyinen manuaalinen datan keruu tekee prosessista riippuvaisen liikennöitsijöistä. Manuaalinen keruu ei ole skaalattavissa suurille datamäärille tai nopeasti muuttuville tilanteille. Manuaalisesti kerätyn datan laatu ja määrä voivat lisäksi olla riittämättömiä päätöksenteon tarpeisiin.

13. Jatkopäätökset ja -ideat

- Millaisia päätöksiä jatkosta on tehty ja millä perustein? (Jatketaan / muutetaan / päätetään)

Päätöksiä jatkon osalta ei ole tehty. Suosittelemme jatkoksi seuraavia toimenpiteitä:

Yhteistyön laajentaminen ja kehittäminen

- laajennetaan yhteistyötä viranomaisten ja liikennöitsijöiden välillä.
- kehitetään molempia hyödyttäviä kilpailutuksia ja sopimusmalleja.

Teknologisten ratkaisujen käyttöönotto ja datan keräyksen kehittäminen:

- otetaan käyttöön ratkaisuja, jotka mahdollistavat paremman datan keräämisen ja analysoinnin vesiliikenteen kontekstissa.
- kokonaistaloudellisesti edullista olisi ottaa yhteiskäyttöön HSL:n olemassa oleva LATI-järjestelmä. Palvelun tuottajilta (liikennöitsijät) toivat esille vahvan tahtotilan olemassaolevan työkalun käyttöön, vs. vastaavan uuden kehittämiseen.