Jari Korpela

Lentoliikenteen aikataulupoikkeamien ennustaminen tekoälyllä

Tietotekniikan pro gradu -tutkielma

10. toukokuuta 2022

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Jari Korpela

Yhteystiedot: jarimatti.korpela@gmail.com

Ohjaajat: Sami Äyrämö ja Joonas Hämäläinen

Työn nimi: Lentoliikenteen aikataulupoikkeamien ennustaminen tekoälyllä

Title in English: Predicting air traffic schedule changes with artificial intelligence

Työ: Pro gradu -tutkielma

Opintosuunta: Ohjelmistotekniikka

Sivumäärä: 90+3

Tiivistelmä: Tutkimuksen tavoitteena oli kehittää ennustemalli, joka ennustaa meno-paluulennon aikataulun vuorokautta ennen lentoa viiden minuutin tarkkuudella. Aikatauluennusteesta nähdään poikkeamat ja myöhästymiset, jolloin sidosryhmille jää aikaa reagoida poikkeavaan aikatauluun. Tutkimusmenetelmänä on suunnittelutiede ja kolmen silmukan malli. Ennuste tehtiin koneoppimisen XGBoost-algoritmilla useasta optimoidusta vaiheesta koottuna kokonaisennnusteena. Vastaavaa ennustemallia ei oltu aiemmin tutkittu. Tutkimuksessa kehitettiin ennustemalli, jolla saavutettiin asetettu tavoite. Opitulla tietämyksellä ja tarkennetulla tavoitteella voidaan tehdä erilaisiin tarpeisiin sopivia ennustemalleja.

Avainsanat: lentoliikenne, aikataulu, ennustaminen, koneoppiminen, tekoäly

Abstract: The aim of the research was to develop a forecasting model that predicts a round-trip flight schedule the day before the flight with an accuracy of five minutes. The schedule forecast indicates deviations and delays, leaving stakeholders time to react to the changed schedule. The research method is design science, and the forecasting method is a step-by-step method of machine learning. The study developed a model to achieve the set goal. A similar prediction model had not been previously studied. With the knowledge learned and the refined goal, forecasting models suitable for unique needs can be made.

Keywords: airtraffic, schedule, predicting, machine learning, artificial intelligence

Esipuhe

Tämä tutkimus on tavallaan kahden elämänpolun yhtymiskohta. Aikanaan valitsin sotilaslentäjän uran jo myönnetyn yliopisto-opintojen sijaan. Ilmavoimien uran jälkeen laitoin jo jalkani tietotekniikan opintojen aloitukseen Jyväskylän yliopistolla, mutta jatkoin kuitenkin liikennelentäjänä seuraavat 14 vuotta. Nyt olen valmistumassa tietotekniikan maisteriksi ja tutkimuksessani hyödynnän kokemustani ilmailun alalta.

Olen saanut keskustella tutkimuksestani lentäjäystävieni, yliopiston ohjaajien, dekaanien ja professorien kanssa. Samalla, kun tutkimus yhdistää kaksi elämänvaihettani, olen päässyt syvällisiin keskusteluihin aiheesta ystävieni kanssa. Heidän elämänkokemuksensa ja erilaiset näkemyksensä ovat täydentäneet omiani ja avanneet silmiäni uusille oivalluksille. Erityisesti kiitän entistä ilmavoimien koelentäjää ja ikuista opiskelijaa Pekka Sievästä, nopeasta ongelmien sisäistämisestä ja kokonaiskuvan puntaroinnista kanssani. Joitain ongelmia on ratkottu syvissä vesissä Päijänteen aalloilla eikä suinkaan tutkijan kammiossa.

Vaimoni Minna saa myös osansa ansioina sitkeästä kannustamisesta ja saattamisessa hyvään lopputulokseen. Tutkimuksen kerronnassa tarvitaan aina maalaisjärkeä ja logiikkaa, jotta työstä tulisi uskottava ja ymmärrettävä. Tavoitehan on, että tutkimus olisi ymmärrettävää ja hyödynnettävää tekstiä sekä toimitusjohtajille että tietotekniikan ammattilaisille.

Jyväskylässä 10. toukokuuta 2022,

Jari Korpela

Termiluettelo

ANN	Artificial Neural Network, neuroverkko -algoritmi
GBDT	Gradient Boosting Decision Trees, satunnainen metsä -algoritmi
DM	Data Mining, tiedonlouhinta
IATA	International Air Transport Association, Kansainvälinen ilmakuljetusliitto
KD	Knowledge Discovery, tietämyksen muodostuksen menetelmä (ks. Fayyad, Piatetsky-Shapiro ja Smyth 1996a)
KDD	Knowledge Discovery in Databases, tietämyksen muodostus tietovarastoista (ks. Piatetsky-Shapiro 1991)
KM	Knowledge Mining, tiedonlouhintaan, tietämyksen muodostukseen ja liiketomintaan liittyvä menetelmä (ks. Äyrämö 2006)
LR	Linear Regression, lineaarinen ennuste -algoritmi
MAE	Mean Absolute Error, keskimääräinen todellinen virhe
MLP	Multilayer perceptron, hermoverkkoarkkitehtuuri
RF	Random Forest, satunnainen metsä -algoritmi (ks. Nykänen 2017)
RVR	Runway visual range, kiitotienäkyvyys
SMOTE	Synthetic Minority Oversampling Technique, datan tasapainotuksen menetelmä (ks. Chawla ym. 2002)
UTC	Coordinated Universal Time, koordinoitu yleisaika
XGBoost	eXtreme Gradient Boosting, gradientilla tehostamisella parannettu satunnainen metsä -algoritmi (ks. Chen ja Guestrin 2016a)

Kuviot

Taulukot

Sisällys

1 Johdanto

Lentoliikenne poikkeaa muista julkisista liikennemuodoista nopeutensa ja yhteysverkostojensa keskittymisen puolesta. Lentoasemaan liittyvät toiminnot ja ympäristötekijät vaikuttavat kokonaisaikaan. Moni tekijä vaikuttaa suunnitellusta aikataulusta poikkeamiseen. Menopaluulento koostuu eri vaiheista ja näitä vaiheita on aiemmin tutkittu, muttei lennon aikataulua kokonaisuutena.

Tutkimusmenetelmänä on suunnittelutiede (Hevner ym. 2004) ja kolmen silmukan malli (Herbert 1996). Ennuste tehtiin koneoppimisen XGBoost-algoritmilla yhdistämällä eri vaiheista koottu kokonaisennnuste. Vaiheissa ennustaminen mahdollisti syötemuuttujien optimoinnin kullekin vaiheelle sekä kunkin vaiheen validoinnin ja korjaamisen tavoitteen mukaiseksi ennen vaiheiden yhdistämistä. Aineistona on Helsingin ja Oulun väliset lennot vuoden pituiselta jaksolta. Myöhästymisten sijaan ennustetaan kunkin vaiheen kesto ja luodaan niistä kokonaisennuste toteutuvalle aikataululle.

Tavoitteena on tutkia, onko tekoälyllä mahdollista ennustaa meno-paluulennon aikataulu vuorokautta ennen lennon lähtöä viiden minuutin tarkkuudella. Tarkkuudeksi asetettiin Finnairin operaatiokeskuksen toiveesta viiden minuutin tarkkuus. Aikataulun ennusteesta ovat kiinnostuneita matkustajat sekä lentoyhtiön, lennonjohdon ja lentoaseman työntekijät. Mikäli ennuste poikkeaa aikataulusta, kullekin sidosryhmälle jää aikaa reagoida poikkeamaan. Yleinen tapa on ilmoittaa lennon poikkeama aikataulusta määräasemalle kyseisen lennon lentoonlähdön jälkeen. Tutkimuksen mukaan aikataulu voidaan ennustaa tarkemmin ja aikaisemmassa vaiheessa kuin nykyiset arviot. Ennuste voidaan muokata eri käyttäjien tarpeiden mukaiseksi.

Seuraavassa luvussa kerrotaan aiheeseen liittyvät taustatiedot, jotka käsittelevät tutkimusmenetelmien, tekoälymenetelmien, säätietojen sekä aikataulun mukaisen lentoliikenteen teorioita ja käytänteitä. Teoriaosuuden jälkeen kerrotaan, kuinka ennustemalli rakentuu. Mallin rakentamisen vaiheet ovat tiedonhankinta, vaiheiden optimointi, laskenta, koostaminen ja lopuksi testaaminen.

2 Tutkimuksen menetelmät ja aihepiirit

Tutkimukseen liittyvät teoriat esitellään tässä luvussa. Ensimmäisessä alaluvussa perustellaan valittu tutkimusmenetelmä ja sen soveltaminen. Toisessa alaluvussa kerrotaan, miten tiedon valintaan liittyvää tiedonlouhintaa käytetään lentotoimintaan liittyen. Seuraavaksi perustellaan valittu tekoälymenetelmä ja sen vaatimukset opetukselle, optimoinnille ja tiedon rikastamiselle. Lopuksi kerrotaan lentotoimintaan ja säätietoihin liittyvät erityispiirteet ja toiminnan periaatteet.

Tavoitteena on ennustemalli, joka on riittävän tarkka siirrettäväksi sellaisenaan käytäntöön. Koko prosessin ajan pyritään ennusteet rakentamaan muuttujilla, jotka ovat tarkkooja ennustaa ja helposti saatavilla. Kokonaisuus on niin laaja, että lopputuloksen onnistuminen vaatii enemmän aiheen kokonaishallintaa kuin yksittäisten menetelmien tarkkuuden optimointia. Menetelmien valintaan ja testaamiseen käytetään IBM SPSS Modeler -ohjelmistoa (*SPSS Modeler* 2022), jolla saadaan nopeutettua tiedon esikäsittelyä ja menetelmän valintaa. Valittu menetelmä opetetaan ja optimoidaan Python SciKit -kirjaston (Pedregosa ym. 2011) avulla. Samalla ohjelmistokoodista saadaan Python-ennustuskomponentti, joka on siirrettävissä tulevan sovelluksen ohjelmistokoodiin. Ohjelmistokoodi on lopullinen tavoite, mutta sitä ennen tarvitaan perusteellista tutkimusta aihepiiristä, jotta ohjelmistokoodi toimisi parhaan tietämyksen mukaisesti.

2.1 Suunnittelutiede

Tutkimusmenetelmäksi valittiin suunnittelutiede, koska sillä kehitetään iteratiivisesti menetelmää vaihe vaiheelta ja ratkaistaan huomattuja ongelmakohtia (Hevner ym. 2004). Kehittämiseen käytetään merkityksellisyyteen ja tietopohjaan liittyviä iteraatioita, jotka lisäävät ympäristön ja tiedon ymmärrystä (Herbert 1996). Työkaluina ovat tietämyksen muodostuksen ja tiedonlouhinnan menetelmät (Fayyad, Piatetsky-Shapiro ja Smyth 1996a). Tiedonlouhinnassa käytetään koneoppimisen satunnainen metsä -algoritmia (Chen ja Guestrin 2016a).

Suunnittelutiede pyrkii luomaan innovaatioita, jotka määrittelevät ideat, käytännöt, tekniset valmiudet, ja tuotteet (Denning 1997). Suunnittelutieteellä luodut ratkaisut ovat harvoin

täysin käyttövalmiita tietojärjestelmiä. Denning (1997) mukaisesti tämä tutkimus pyrkii luomaan toteuttamiskelpoisen mallin, joka on ideoitu, rakennettu, testattu, korjattu ja on valmiiksi toteuttamiskelpoinen malli sovellukselle. Malli toimii kyseisessä ympäristössä ja on laajennettavissa muuallekin, kuten tässäkin tutkimuksessa myöhemmin todetaan.

Lisäksi mallia voidaan muokata ympäristön muuttuessa, kuten Johansson, March ja Naumann (2003) totesivat. Menetelmä sopii erityisesti silloin, kun luodaan uusi aiemmin toteuttamaton malli. Menetelmällä analysoidaan, suunnitellaan, toteutetaan ja hallitaan toteutettavaa mallia niin, että tuleva tietojärjestelmä voidaan toteuttaa ja käyttää tehokkaasti (Tsichritzis 1997).

Ennustemallin kehittämiseen käytetään rakenna ja arvioi -menetelmää. Rakenna ja arvioi -silmukassa arvioidaan mallia vertaamalla tuloksia todellisiin tapahtumiin ja parantaen mallia jokaisella iteraatiolla. Iteraatioissa parannetaan ennusteita, mallin rakenteita, arkkitehtuuria ja tuloksen esitystä. (March ja Smith 1995)

Mallin kehittämisessä noudatetaan seitsemän kohdan listaa, jonka alunperin julkaisi Klein ja Myers (1999). Lista noudattaa suunnittelutieteen vakiintuneita ja parhaita käytänteitä. Seitsemän kohtaa ja niiden soveltaminen on kerrottu taulukossa 1. (Hevner ym. 2004)

Vastaavaa kokonaisennustetta ei oltu aiemmin tutkittu, joten tutkimus kehittää ratkaisun lentotoiminnan ja siihen liittyvien aikataulun mukaisten liiketoimintojen optimointiin. Tarve ennustemallille oli määritetty lentoyhtiössä ja haluttiin ratkaisua todelliseen ongelmaan. Tutkimuksen tavoitteena on selvittää, onko mahdollista luoda toimiva malli aikataulupoikkeamien ennustamiseen. Mallia parannettaisiin tutkimuksessa saadun tietämyksen perusteella, mikäli ensimmäinen iteraatio täyttäisi tavoitteen vaatimuksen. Tällaiseen kehittämiseen sopii hyvin suunnittelutieteen iteratiivisuuden periaate. Iteratiivisuutta on mallintanut Hevner (2007) kolmella silmukalla, joita kuvataan seuraavissa kappaleissa. Silmukat sopivat hyvin myös myöhemmin esiteltävään KM-menetelmään (Äyrämö 2006).

Kuviossa 1 on kolmen silmukan malli, jonka ensimmäinen silmukka on merkityksellisyyssilmukka. Se on ympäristön vaatimusten ja tutkimusmenetelmien välinen vuorovaikutus, jossa liiketoiminnan tarpeita ja mahdollisuuksia arvioidaan. Tämän silmukan ympäristöön kuuluvat lentotoiminnan sidosryhmien teknologiat ja toiminnot vaatimuksineen. Tästä sil-

Suunnitteluvaihe	Kuvaus tässä tutkimuksessa
Suunnittele artefaktiksi	Tutkimuksen täytyy tuottaa artefakti, joka on malli toteutettavalle ennustus-sovellukselle.
Ongelman merkityksellisyys	Tutkimuksen tavoitteena on kehittää ratkaisu lentotoiminnan ja siihen liittyvien aikataulun mukaisten liiketoimintojen optimointiin.
Suunnitelman arviointi	Tulos pitää todentaa oikeasta ympäristöstä saadulla datalla. Hyödyllisyys, laatu ja tehokkuus vahvistetaan tietojoukosta erotetulla testidatalla.
Tutkimuksen vaikutukset	Toteutus tulee olla ratkaisu aikataulupoikkeamien ennustamiseen. Tutkimuksen tulee parantaa todennettavasti aikatauluennusteen tarkkuutta sekä ennusteen aikaväliä.
Tutkimuksen täsmällisyys	Tutkimus perustuu kuvatun suunnittelutieteen menetelmien ja prosessien käyttöön mallin luomisen ja arvioinnin yhteydessä.
Suunnittelu etsimisprosessina	Tehokkaan mallin etsiminen vaatii aiempien tutkimusten hyväksikäyttöä samalla huomioiden tutkimuksen avulla saatua uutta tietämystä.
Tutkimuksen kommunikointi	Tutkimuksen kuvauksessa huomioidaan sekä teknologisesti orientoituneet että johtamisorientoituneet lukijat.

Taulukko 1. Suunnittelutieteen ohjeet

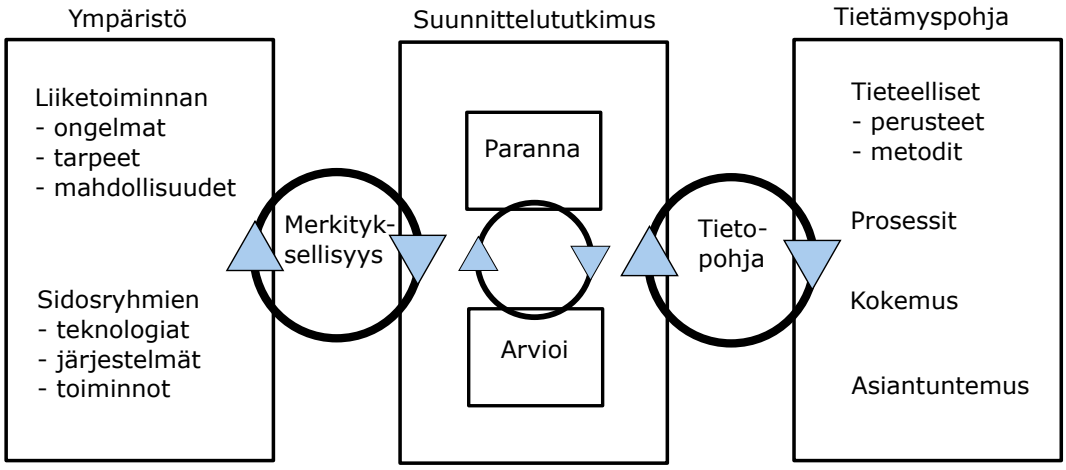
mukasta saadaan liiketoiminnan ongelma ja tutkimuksen tavoite. Koska tutkimuksessa tutkitaan liiketoiminnan todellista ongelmaa, tutkimuksesta on mahdollista tulla merkityksellinen ja siten saavuttaa tuloksilla käytännön hyötyä. (Hevner 2007)

Toinen silmukoista on tietopohja-silmukka. Se pyrkii huomioimaan tieteellisten perusteiden, kokemuksen ja asiantuntemuksen vaikutuksia menetelmään. Tätä silmukkaa varten on kerätty soveltuvia tutkimuksia aihepiiristä. Kokemus ja asiantuntijuus täydentävät lopulta teo-

rioiden ja menetelmien perusteella olevaa tietopohjaa, jota silmukassa mahdollisesti löydetty uusi tietämys edelleen täydentää. (Hevner 2007)

Kolmannessa silmukassa eli tutkimus-silmukassa näitä ympäristön ja tietämyspohjan vaatimuksia sovelletaan ja hyödynnetään käyttäen hyväksi valitun tutkimusmenetelmän arvioi ja paranna -silmukkaa. Näin lentotoiminnan vaatimukset ja tietämyspohja ohjaavat tutkimusta ja tutkimus luo uutta tietämystä ja rikastaa tietopohjaa. Kirjallisuuskatsauksen ja asiantuntija-arvioiden avulla kerättyä tietoa hyödynnetään menetelmien ja syötemuuttujien valinnoissa. (Hevner 2007)

Tutkimusmenetelmien tulokset luovat uusia tarpeita rikastaa tietämyspohjaa ja teorioita (Herbert 1996). Näin rakentuu kolmen silmukan välinen kokonaisuus, joka parantaa mallia kohti tutkimuksen tavoitetta. Mallia arvioidaan seuraamalla sen ennusteita toteutuneisiin arvoihin. Vaiheiden laskentaa kuvaavassa luvussa kerrotaan tarkemmin, kuinka tutkimusmenetelmän silmukoita käytetään. (Hevner 2007)



Kuvio 1. Kolmen silmukan malli (Hevner 2007)

Merkityksellisyys-silmukalla pyritään löytämään toimintaympäristölle oleelliset asiat tutkimuksen tavoitteen kannalta. Tietopohja-silmukka puolestaan lisää ymmärrystä parhaan tiedon valinnalle. Tietämyksen muodostuksella pyritään selvittämään näitä molempia (Hevner 2007). Seuraavassa alaluvussa esitellään tietämyksen muodostuksen menetelmiä. Käytetyt menetelmät vastaavat silmukoiden toimintaa.

2.2 Tietämyksen muodostus -menetelmiä

Tiedonlouhintaa (engl. DM) laajempi menetelmä on tietämyksen muodostus (engl. KD). Tietämyksen muodostuksesta on menetelmää edelleen kehitetty tietämyksen muodostukseksi tietovarastoista (KDD), joka julkaistiin ensimmäisen kerran 1989 (Piatetsky-Shapiro 1991).

Dunham (2003) erotteli tiedonlouhinnan ja tietämyksen muodostuksen tietovarastoista seuraavasti: "Tietämyksen muodostus tietovarastoista on prosessi, jolla pyritään löytämään hyödyllistä tietämystä ja malleja tiedosta. Tiedonlouhinta on algoritmien käyttöä niin, että erotellaan tiedosta tietämys ja mallit, jotka ovat löydetty tietämyksen muodostuksella."

Myöhemmin Fayyad, Piatetsky-Shapiro ja Smyth (1996a) totesi artikkelissaan tietämyksen muodostuksen olevan merkityksellinen, kun tiedon määrä alkoi lisääntymään moninkertaiseksi. Tämän tutkimuksen tietomäärä ei vaadi KDD-menetelmää, mutta menetelmän prosessi on sopiva tähän tutkimukseen.

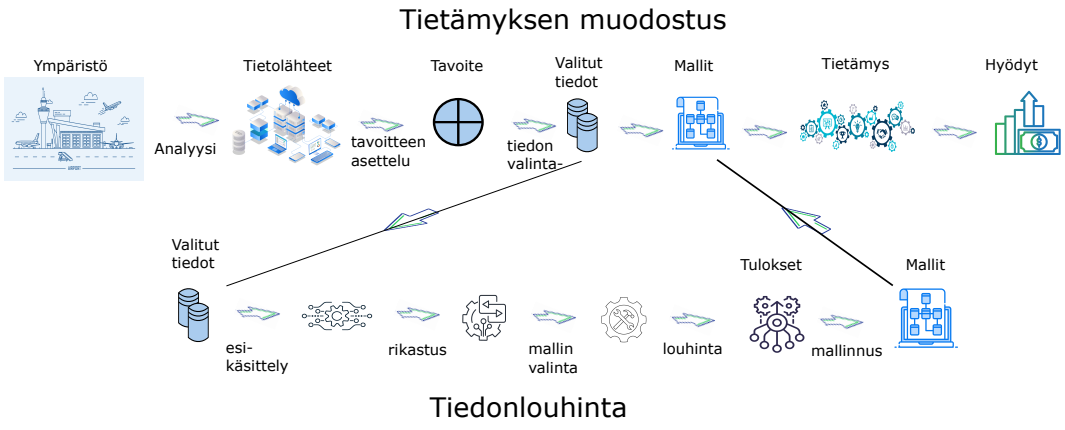
Äyrämö (2006) on väitöskirjassaan kehittänyt edelleen kokonaisuutta tietämyksen muodostukseksi (KM), joka sisältää tiedon louhinnan lisäksi tietämyksen muodostuksen osat sekä laajemman kokonaisuuden liiketoimintaan liittyen. Hänen esittämänsä malli tukee myös esitettyä kolmen silmukan mallia. Tämä yhdistetty tietämyksen muodostus sisältää kuvan 2 mukaisesti alla kerrotut vaiheet, joista pääosat Fayyad, Piatetsky-Shapiro ja Smyth (1996b) esittivät mallissaan. Tiedon ja tietämyksen erottelu vastaa valitussa suunnittelututieteessä olevia tietopohja- ja merkityksellisyys-silmukoita. Tietopohja on yhteydessä tiedonlouhintaan ja merkityksellisuuden on yhteydessä tietämyksen muodostukseen.

Tietämyksen muodostuksessa (KM) ovat seuraavat vaiheet, joissa edetään ja voidaan iteratiivisesti palata aiempiin vaiheisiin (Äyrämö 2006). Alla olevan listan vaiheet ovat kuvassa 2.

- Toimialueen analyysi: Lisää ymmärrystä aiheesta. Selvittämällä tiedon omistajat, alkuperä, tarpeellisuus ja muoto, saadaan tieto tarvittavista resursseista. Resursseihin vaikuttavat, mitä tietoja tarvitaan ja paljonko tiedon hankkiminen vie aikaa.
- Tavoitteen asettelu: Tiedostetaan, mistä tarvittava tieto on saatavilla sekä kuka sen voi antaa ja koska se on saatavilla. Tiedon saatavuudelle pitää asettaa tavoite suhteessa

tiedon tarpeeseen.

- Tiedon valinta: Teknisesti on selvitetty, miten tieto saadaan käyttöön. Perustuen tietämykseen valitun tiedon hyödyllisyydestä, valitaan syötemuuttujat. Tärkeimmät tiedot valitaan riippumatta tiedon muodosta. Esikäsitellään tieto sopivaksi huomioiden käytössä olevat resurssit (henkilöstö, laitteisto ja aika).
- Tiedonlouhinta valituilla syötemuuttujilla: Saadaan mallit ja grafiikat tuloksista. Tiedon käsittely ja menetelmät eivät näy toimialueen asiantuntijoille. Tiedon käsittely voidaan tehdä sopivimmalla valitulla menetelmällä tuntematta toimialuetta. Tehtävät voidaan jakaa oman alansa asiantuntijoille.
- Tulkinta ja arviointi: Lisää tietämystä ennalta määritettyyn tavoitteeseen. Lisääntynyt tietämys parantaa liiketoiminnan ymmärrystä ja tietojen välisiä suhteita ja lisää taustatietoa. Tärkeätä tietoa saadaan tutustumalla toimintaympäristön piilossa olevaan ja hiljaiseen tietoon. Tämä tieto jää helposti hyödyntämättä ja tiedon puute aiheuttaa epätarkkuutta aineistossa. Tällaista tietämyksen hallintaa kuvasivat Kärkkäinen ym. (2003) osaraporteissaan.
- Hyödyntämisen tavoitteena on, että saadut tulokset johtavat toiminnan paranemiseen.



Kuvio 2. KM-malli, mukaellen Sami Äyrämö (2006)

Prosessissa kuvattu tiedonlouhinta voidaan tehdä itsenäisesti ja se vaatii syvempää teknistä ja laskennallista osaamista. Sen vaiheet ja yksityiskohdat voidaankin usein piilottaa toimialan asiantuntijoilta, vaikka tiedonlouhinnan ratkaisut algoritmeissa ovatkin olennaisen tärkeitä lopputuloksen kannalta. (Äyrämö 2006) Näin asiantuntijoiden on mahdollista toimia molemmissa silmukoissa omilla vahvuusalueillaan itsenäisesti. Itsenäisesti tiedonlouhinnassa

toimittaessa on varmistettava, että saatu tietämys siirtyy paranna ja arvioi -silmukkaan.

Tiedonlouhinnan osat kuvassa 2 ovat (Äyrämö 2006):

- Esikäsittely: Tieto käsitellään tavoitteen mukaisesti puuttuvien ja virheellisten tietojen osalta.
- Tiedon muuttaminen: Tieto muutetaan niin, että lukuväli, muoto ja tyyppi ovat oikeita käytetylle menetelmälle.
- Menetelmän valinta: Oikea menetelmä tiedon käsittelyä varten valitaan. Valintaan vaikuttavien syöte- ja vastemuuttujien tyypit sekä tutkimuksen tavoite.
- Laskenta: Valittu menetelmä optimoidaan lähtötiedon ja tuloksen saavuttamiseksi.
- Esitys: Saadut tulokset esitetään muodossa, jossa ne ovat hyödynnettävissä tavoitteen mukaisesti.

Tiedonlouhinnassa käytetään tietoon ja tavoitteeseen sopivaa menetelmää. Valinnassa on tiettyjä perusasioita, mutta jäljelle jäävät menetelmät pitää testata ja valita niistä paras. Tässä tutkimuksessa tehdään kokonaisennuste useista lennon vaiheiden ennusteista ja tavoitteena on käyttää kaikissa ennusteissa samanlaista algoritmia. Tällä pyritään ajan ja työn säästöön. Aiemmat tutkimukset, nopeat testausmenetelmät ja menetelmien ymmärtäminen auttavat valinnassa.

Tietämyksen muodostus ja tiedonlouhinta sisältävät erilaisia tehtäviä, joiden suhteellista tärkeyttä ja aikavaatimusta voidaan vertailla. Pyle (1999) vertaili tietämyksen muodostuksen, tiedonlouhinnan ja mallinnuksen suhteita. Hänen jaottelunsa sisältää samat tehtävät, kuin aiemmin on esitelty kuvassa 2. Taulukossa 2 on hänen erittelynsä ajankäytön ja tärkeyden suhteen eri tehtävissä.

Tuo jaottelu kuvaa, miten tärkeä tietämyksen muodostus on ja paljonko siihen yleensä käytetään aikaa. Osuudet vaihtelevat datan ja tutkimustavan mukaan. Ramírez-Gallego ym. (2017) kertovat hyvin, kuinka näihin osuuksiin vaikuttavat tiedon eheys, määrä, rakenne ja uudet laskentamenetelmät. Heidän arvionsa tiedon esikäsittelyn aikavaatimuksesta oli > 50 % tutkimusajasta.

Näillä perusteilla tiedonlouhinnan esikäsittely vaatii 50–75 % kokonaisajasta. Tietämyksen

Tehtävä	Aikavaatimus	Tärkeys
Ongelman määrittely	10	15
Ratkaisun selvittäminen	9	14
Toteutuksen määrittely	1	51
Tiedon esikäsittely	60	15
Tiedon tutkiminen	15	3
Tiedon mallinnus	5	2

Taulukko 2. Tehtävien aikavaatimukset ja tärkeydet, (Pyle 1999)

muodostuksen tärkeys korostuu, koska siinä tehdään pohjatyö tiedonlouhinnalle. Seuraavan alaluvun tekoälymenetelmät ovat suoraviivaisia ja nopeita, kunhan tiedonlouhinta on hyvin tehty. Ajallisesti tekoälymenetelmän mallinnus vie noin 5 % ajasta eikä ole kokonaisuuden kannalta tärkeydessä kuin noin 2 % (Pyle 1999). Tiedon mallinnus vaatii kuitenkin hyvää menetelmätuntemusta ja tehokasta laskentaa.

2.3 Tekoälymenetelmiä

Tässä luvussa esitellään tähän tutkimukseen liittyvää teoriaa tekoälymenetelmistä. Tekoälyn tyypit voidaan jaotella monella tavalla. Toiminnallista jaottelua on kuvattu tutkielmassa , joka kertoo lentoyhtiön tekoälyn hyödyntämisestä. Siinä kuvataan, että tekoälyn kyvykkyydet ovat verrattavissa ihmisen aisteihin ja toimintaan. Kuvantunnistus vastaa näköaistia. Puhe tekstiksi -kyky vastaa kuuloaistia. Teksti puheeksi -kyky vastaa ihmisen puhumista. Tekstianalyysit vastaavat luetun ymmärtämistä. Tiedonkäsittely vastaa ihmisen pikamuistia ja pitkää muistia. Koneoppiminen sisältää sekä oppimisen että päättelyn. Päättelyä ja oppimista tarvitaan tässä tutkimuksessa. (Korpela 2019)

Yleisesti koneoppiminen on jaoteltu toimintansa perusteella ohjattuun oppimiseen, ohjaamattomaan oppimiseen, vahvistusoppimiseen. On olemassa muitakin tyyppejä, kuten opitusta oppiminen. Ohjatussa oppimisessa opetusdatassa annetaan mukana vastemuuttuja, joka on haluttu oikea tulos. Ohjaamattomassa oppimisessa ei tavoitella lähdemuuttujan arvoa, vaan esitetään suhteita, riippuvuuksia ja samanlaisuuksia aineistosta. Vahvistetussa oppimi-

sessa kohdetieto korvataan ympäristön antamalla palautteella. Opitusta oppimisessa opitaan algoritmin omista kokemuksista. (Zhang 2010)

Käytetyt algoritmit valitaan perustyypin, syötemuuttujien ja halutun tuloksen eli vastemuuttujan mukaan. Syötemuuttujat voivat olla jatkuvia lukuja, epäjatkuvia lukuja tai luokkatietoja. Vastemuuttuja on tämän tutkimuksen aikaennusteissa jatkuvaa tietoa. Jatkuvan tiedon ennustamiseen voidaan käyttää regressio-analyysiä. Regressioanalyysissä tutkitaan yhden tai useamman syötemuuttujan yhteyttä valittuun vastemuuttujaan eli haluttuun tulokseen. Yleensä käytetään lineaarista regressiota, mutta on olemassa myös epälineaarinen regressio. (Darlington ja Hayes 2017)

Tekoäly-algoritmeihin on tarvittaessa rakennettu valmiiksi regressiolaskenta, jota algoritmin virhefunktiolla optimoidaan pienimmän virheen saamiseksi koko aineistossa. Erityyppisissä ennusteissa virhefunktio on pyritty sovittamaan ennustemallin mukaiseksi, jotta menetelmän sisäisten muuttujien avulla algoritmista saadaan pienin kokonaisvirhe. Tälle kokonaisvirheen funktiolle haetaan minimiarvo derivoimalla, jolloin saadaan algoritmin sisäisille muuttujille parhaat arvot kokonaisvirheen minimoimiseksi. Virhefunktion minimoiminen vaatii aina käytetyn algoritmin mukaisen virhefunktion. Lentotoiminnan aikoja ennustettaessa tiedetään historiatiedoista toteutunut aika, joka on vastemuuttuja. Valitulle algoritmille opetetaan sisäiset muuttujat, jotka antavat annetuilla syötemuuttujilla pienimmän virheen suhteessa vastemuuttujaan.

Tullaan siis tarvitsemaan ohjatun oppimisen menetelmä, joka on käsiteltävään tietoon sopiva koneoppimisen menetelmä. Tutkimuksessa ennustaminen tulee perustumaan historialliseen tilastotietoon eikä sääntöpohjaiseen ohjelmointiin. Riittävällä määrällä opetusdataa löytyy algoritmi, joka antaa myös uusilla tapauksilla mahdollisimman oikean ennusteen, mitä tulisi tapahtumaan. Sääntöpohjaiset ohjelmat korvataan tilastotieteen ja uusien virheen minimointitapojen avulla. (Weibo ym. 2017)

Koneoppiminen kykenee laskemaan monen tekijän riippuvuussuhteista mallin, joka antaa pienimmän keskimääräisen virheen koko tietojoukolle. Liian monta syötemuuttujaa aiheuttaa kuitenkin moniulotteisuuden ongelman. Esimerkiksi käyrän kuvaus kaksiulotteisena voi vaatia kymmenen pistettä ja kuvaaja kolmiulotteisena pintana kolmella syötemuuttujalla vas-

taavasti sata pistettä. Neljännen ulottuvuuden lisäys vaatisi tällä oletuksella 10 000 pistettä ja viides ulottuvuus edelleen 100 000 pistettä. Ulottuvuuden kasvaessa tarvittavan tiedon määrä kasvaa nopeasti suureksi. Tällöin mallinnukseen tarvittavan tiedon määrä kasvaa nopeasti liian suureksi (Verleysen ja Francois 2005). Keskenään korreloivat syöteparit tulisi yhdistää (Yu ja Liu 2003). Turhien muuttujien karsiminen vähentää moniulotteisuutta ja vaadittavien syötemuuttujien määrää vähenee (Kanevski 2009).

Aiempien lentoliikenteen myöhästymisiin liittyvien tutkimusten mukaan useimmin käytetty algoritmi oli satunnainen metsä -algoritmi. Myös SPSS Modeler valitsi kyseisen algoritmin eri vaihtoehdoista. Tässä tutkimuksessa valittiin näillä perusteilla käytettäväksi satunnainen metsä -menetelmä, joka koostuu useista päätöspuista.

Päätöspuut käsittelevät dataa jakamalla tietoa puun haaroissa osiin ja osia edelleen haaroittamalla. Haaroitukseen käytetään erilaisia algoritmeja kuten ID3, C4.5, CHAID ja MARS. Kyseisillä algoritmeilla tieto jaetaan haaroihin ja jaon hyvyyttä optimoidaan esimerkiksi gini- tai entropia-lisää parantamalla. Näin valitaan parhaat haaroittavat syötemuuttujat ja paras haaroittava arvo. Molemmilla arvoilla voidaan laskea jaon hyvyys eli informaatiolisä. (Bell 2020)

Tutkimuksessa syötemuuttujien merkitystä ennusteeseen lasketaan permutaation tärkeydellä. Permutaatio tärkeyttä lasketaan sekoittamalla menetelmän kunkin syötemuuttujan arvoja vuorollaan satunnaisesti ja vertaamalla, paljonko kunkin piirteen sekoittaminen vaikuttaa ennusteen tulokseen. Permutaatio-kuvaajasta näkee, paljonko kunkin syötemuuttujan sekoittamisen vaikutus on lopputuloksessa. (Breiman 2001)

Tämän arvioinnin vuoksi on myös tärkeää karsia keskenään korreloivia muuttujia datasta. Kun tärkeän piirteen arvoja permutoidaan, sen kanssa korreloiva toinen syötemuuttuja korjaa ennusteen tulosta oikeaksi. Tällöin saadaan väärää tietoa, että kyseisen syötemuuttujan muuttamisella oli vähäinen vaikutus tulokseen ja syötemuuttuja todetaan merkitykseltään vähäiseksi. Mikäli ennuste toimii hyvin, mutta permutaatio antaa tasaisesti pieniä arvoja, syötemuuttujissa on vielä keskenään korreloivia arvoja. (Gregorutti, Michel ja Saint-Pierre 2017)

Päätöspuista on kehitetty satunnainen metsä -menetelmä. Satunnaisessa metsässä eri päätös-

puut tukevat toistensa heikkouksia ja muodostavat yhden vahvemman mallin (Polikar 2012). Päätöspuiden erilaisuus saadaan aikaan jättämällä eri puista tietyt tai satunnaiset syötemuuttujat pois (Breiman 2001). Käyttämällä keskenään erilaisia päätöspuita saadaan mallista hyvin yleistävä ja vältetään ylioppimista. Satunnaisessa metsässä voidaan eri puille käyttää erilaisia laskentamenetelmiä, millä voidaan tehostaa heikompien alueiden ennustavuutta. Tällaista tehostusta kutsutaan skaalautuvaksi tehostamiseksi (Chen ja Guestrin 2016b).

XGBoost on tehostettu, gradientti ja skaalautuva satunnainen metsä -menetelmä, jonka esitteli Chen ja Guestrin (2016a). Heidän mukaansa XGBoost perustuu skaalautuvaan tehostettuun algoritmiin ja kuuluu kokoonpanomenetelmiin (engl. ensemble method). Tehostaminen viittaa tekniikkaan, jossa luodaan useita malleja peräkkäin ja jokainen malli yrittää parantaa edellisen mallin tarkkuutta. Perustana on päätöspuista koostuva ennuste, jota voidaan käyttää luokitteluun sekä regressioon (Mitchell ja Frank 2017). XGBoost on tehokas ennustaja kokoonpanomenetelmänsä ja tehostetun gradientti skaalauksen vuoksi. Menetelmä mahdollistaa laskentaprosessorin moniydinlaskennan (Chen ja Guestrin 2016a).

Ero perinteisellä satunnaisella metsällä ja tehostetulla skaalautuvalla menetelmällä on siinä, että perinteinen satunnainen metsä -menetelmä ennustaa syötemuuttujista puiden ennustettua vastemuuttujaa, kun taas tehostettu skaalautuva menetelmä ennustaa vaihe vaiheelta tarkentuvaa tulosta, joka saadaan edellisten $(n-1)$ puiden tuloksesta ja seuraavan ennustettavan (n) syötemuuttujan muutoksen (Δ_n) ennustamisesta. Ennustettu muutos kerrotaan korjauskertoimella η. Iteraatioiden lisääntyessä η pienenee. Jotta oppiminen ei menisi liian tarkaksi opetusdatalle, käytetään sakotuskerrointa, joka kasvaa liian pitkälle viedyssä opetuksessa. Lisäksi sakotustermin on oltava helposti derivoituva tappiofunktion laskemiseksi (Mitchell ja Frank 2017). Tehostetulla skaalautuvalla menetelmällä edellisen ennusteen virhe annetaan aina vastetiedoksi seuraavalle ennusteelle.

$$F_n = F_{n-1} + \eta . \Delta_n \tag{2.1}$$

Virhefunktio on XGBoost-algoritmin ydin, jonka esitteli Chen ja Guestrin (2016a) sekä myöhemmin Zhang ym. (2018). Se sisältää tappiofunktion l sekä sakotustermin Ω. Tappiofunktio l laskee opetusdatan $(n-1)$ kokonaisvirhettä ja sakotustermi Ω pyrkii estämään mallin

monimutkaisuuden rankaisten liian yksityiskohtiin menevää mallia.

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i^t) + \sum_{k=1}^{K} \Omega(f_k) \tag{2.2}$$

Laskennassa käytetään satunnaisen metsän puita yksi toisensa jälkeen. Jokaisen puun jakautumiset ja karsinnat optimoidaan ja syötemuuttujien valinta satunnaistetaan. Menetelmässä voidaan optimoida kokonaisuutta alkaen päätöspuiden, satunnaisen metsän, gradient laskeutumisen ja tehostuksen ominaisuuksia muuttamalla. Nämä ominaisuudet muokataan syötemuuttujien kannalta optimeiksi, jotta kokonaisvirhe olisi pienempi eli opetus toimisi parhaiten. Zhang ym. (2018)

Näitä ominaisuuksia voidaan muokata antamalla niiden tiedot kokoelmassa algoritmin sisäisiä muuttujia eli hyperparametrejä. Jokaisella algoritmilla on omat muuttuvat ominaisuutensa, joita muokataan hyperparametrien avulla. Näiden hyperparametrien arvojen valinta optimoidaan tekemällä erilaisten parametrikombinaatioiden vertailu automaattiseksi. Moniulotteisesta hyperparametrien yhdistelmästä luodaan verkko, jonka kaikki kombinaatiot käydään läpi ja lasketaan paras ennustajakombinaatio. Tämän ennustajan hyperparametrien alueelle luodaan uusi paikallisempi verkko ja jatketaan näin optimointia, kunnes löydetään hyperparametrien paras yhdistelmä algoritmin parhaalle ennustavuudelle. Tässä tutkimuksessa käytettiin Random Grid Search -etsintää, joka hakee satunnaisesti moniulotteisesta tiedosta aluetta, joka voisi olla pienimmän virheen aluetta. Näin vältytään käymästä läpi kaikkia pisteitä moniulotteisessa koordinaatistossa. Satunnaisesti hyperparametrejä etsivä menetelmä etsii kyseiselle aineistolle viritetyn laskentamallin, joka antaa pienimmän virheen. Tätä satunnaista virittämistäkin ohjataan omilla hyperparametreillä. (Bergstra ja Yoshua 2012)

Opetuksen tulisi tuottaa malli, joka on yleistettävissä vastaavalle hieman erilaiselle tiedolle. Jos mallin opetusta jatketaan liian pitkälle, malli oppii kyseiselle tiedolle aina vain tarkemman ennusteen. Samalla muulle vastaavalle tiedolle ennustetarkkuus huononee. Mallin tarkkuutta tulee opettaa siihen asti, että malli ei ala antaa virheellistä ennustetta muulle vertailutiedolle. Yleistyvyyttä parantamaan on menetelmiä, joilla opetus voidaan keskeyttää, kun aletaan havaita ylioppimista. K-ositettua ristiinvalidointia käytetään opetuksen ohessa estämään ylioppimista. Menetelmässä opetusdata jaetaan K-määrään kansioita. Jokainen kansio

toimii vuorollaan validoinnin pienenä testikansiona. Testikansio tarkkailee, ettei opetuksessa tapahdu ylioppimista. Ylioppiminen havaitaan näiden kansioiden validointivirheestä. Kymmenellä kansiolla saadaan opetusdatalla jo ennuste, joka ei todennäköisesti ole ylioppinutta testidatallakaan. (Haykin 2009)

Validointi voidaan tehdä myös erillisellä validointidatalla, mutta tutkimuksessa on päätetty käyttää vain testidataa ja ristiinvalidointia. Näin aineistosta saadaan enemmän tietoa opetusdatalle. Validointivirheen mittauksen lisäksi kannattaa jo tutkimuksen alkuvaiheessa olla valittuna testidata. Testidata pidetään sivussa opetukselta. Kun eri menetelmillä on opetettu lopullinen malli, lopuksi käytetään testidataa tehdylle mallille syötemuuttujina. Testidatalla ennustamalla katsotaan, miten hyvin malli ennustaa ennen näkemättömällä tiedolla. Tämän lopullisen testidatan tarkkuutta mitataan testivirheellä. (Bishop 2006)

Opetus- ja testivirheet ilmoitetaan tässä tutkimuksessa keskimääräisenä absoluuttisena virheenä. Tilastotieteessä se on havaintoparien keskimääräinen ero mittayksiköllä. Tässä tutkimuksessa MAE-arvo saadaan summaamalla ennusteen arvon y ja todellisen arvon x erotuksen itseisarvot, jonka jälkeen kyseinen summa jaetaan mittausten lukumäärällä. Se kuvastaa, paljonko ennusteet keskimäärin eroavat todellisista arvoista.

$$MAE = \frac{\sum_{i=1}^{n} | y_i - x_i |}{n} \tag{2.3}$$

Ennusteiden vastaavuutta todellisiin arvoihin voidaan mitata korrelaatiolla. Tämän tutkimuksen kuvaajissa on käytetty Pearsonin korrelaatiota. Se on yleisesti käytetty mittari lineaarisesti suhteutuvien arvoparien vertailussa. Psykologian, politiikan ja biotieteen tutkimuksissa vastaavuutta on kuvattu sanallisesti vahvaksi tai erittäin vahvaksi, kun arvo on 0,7 tai yli. Arvolla 0,6 korrelaatio kuvattiin kohtalaiseksi tai vahvaksi vastaavuudeksi. (Akoglu 2018)

Algoritmin optimointi kuuluu tiedonlouhinta-vaiheeseen, joka on suunnittelutieteen tietopohjasilmukan aluetta, vaikkakin se hyödyntää myös merkityksellisyys-silmukkaa. Tiedonlouhinnassa käytetyt tiedot kuuluvat vastaavasti tietämyksen muodostukseen ja merkityksellisyyssilmukkaan. Tutkimuksessa tiedonlouhintaan valittu menetelmiä perustellaan aiempien tutkimusten tuloksilla. Seuraavassa alaluvussa kerrotaan aiemmista tutkimuksista, mitä mene-

telmiä oli todettu hyviksi ja millaisia syötemuuttujia oli käytetty.

2.4 Aiemmat tutkimukset

Lentojen myöhästymisistä on tehty paljon tutkimuksia, mutta ne ovat rajattu yhteen lennon vaiheeseen ja myöhästymisen ennustamiseen. Myöhästymistä on ennustettu tapahtumana tai myöhästymisen minuuttimääränä. Yhdistettyä mallia eli eri ennusteista luotua kokonaisuutta ei ole tutkittu. Aiempien tutkimusten tuloksia voidaan hyödyntää meno-paluulennon aikataulun ennustamisessa. Aiemmissa tutkimuksissa on keskitytty parhaan ennustemenetelmän valintaan, mutta tässä tutkimuksessa käytetään aiemmin hyväksi todettuja menetelmiä. Eri vaiheiden tutkimuksia hyödyntämällä valitaan yksi yhteinen koneoppimisen menetelmä ja optimoidaan syötemuuttujat.

Alla on lueteltu eri lennon vaiheisiin liittyvät tutkimukset ja niissä todetut parhaat menetelmät ja syötemuuttujat. Kaikkia syötemuuttujia on testattu tässä tutukimuksessa, mutta vain osa on jäänyt ensimmäisen iteraation jälkeen syötemuuttujiksi.

Taulukossa 3 ovat poikkeamaennusteet tapahtumina. Usein myöhästyminen oli luokiteltu eri aikakynnyksittäin tapahtumaksi. Tasapainotuksessa käytettiin SMOTE-tekniikkaa kasvattamaan pienemmän luokan esiintymiä. Viidestä tutkimuksesta neljä käytti satunnainen metsä -menetelmää. Näistä tutkimuksista ei voida tehdä vertailua ennusteen tarkkuudesta, koska nämä ennusteet olivat luokittelevia.

Tutkimus	Menetelmä	Syötteet
(Pamplona ym. 2018)	ANN	viikonpäivä, lentoyhtiö, tuloaika ja lähtöaika
(Choi ym. 2016)	RF (SMOTE)	lähtö- ja määräkentän sää- ja ennustetiedot , aikataulutiedot
(Chen ja Li 2019)	RF	aikataulutiedot, säätiedot
(Horiguchi ym. 2017)	XGBoosting	lento- ja matkustajatiedot
(Thiagarajan ym. 2017)	XGBoosting ja RF	aikataulu- ja säätiedot

Taulukko 3. Myöhästymiset yleensä

Taulukossa 4 ovat poikkeamien ennusteet ennen portilta lähtöä. Tähän taulukkoon on lis-

tattu viisi ennusteita, jotka koskivat sekä lähtö- että tuloaikoja. Taulukoissa ovat vain syötemuuttujat, jotka sopivat tässä käsiteltävään kiinteään kenttäväliin. Kolmessa tutkimuksessa käytettiin satunnainen metsä -menetelmää ja kolmessa neuroverkko-menetelmää. Kolmessa tutkimuksessa käytettiin syötemuuttujina matkustajien lukumäärää ja yhdessä matkustajamäärä oli korvattu istuinpaikkojen lukumäärällä.

Tutkimus	Menetelmä	Syötteet
(Halmesaari 2020)	XGBoost	matkustajien lukumäärä, matkatavaran määrä, kuukausi, kellonaika, erikoismatkustajien määrä
(Hassel 2019)	MLP	tuloaika portille, jäävät matkustajat, tulevat matkustajat, erikoismatkustajien määrä, seisontapaikka, matkatavaran määrä, tankkaus
(Khanmohammadi, Tutun ja Kucuk 2016)	ANN	tulokenttä, päivämäärä, viikonpäivä, lähtöaika portilta, tuloaika portille, tulon myöhästyminen, myöhästymisen syyt
(Laurence 2021)	RF, ANN	konetyyppi, terminaali, istumapaikkojen määrä, liikennemäärä, kuukausi, vuorokauden minuutit, vuoden päivät, viikon päivät, lämpötila, kosteus, lämpötila yli 3C, puuskat, tuulen suunta, sivutuuli, tuuli yli 30 kts
(Guo ym. 2021)	RFCA, RF	istumapaikkojen määrä, suunniteltu kääntöaika, edellinen lento, liikennemäärä, lentokentän matkustajien lukumäärä

Taulukko 4. Kääntöajan ennusteet

Taulukossa 5 ovat poikkeamien ennusteet koneen rullausajalle. Kaikissa kolmessa oli käytetty satunnainen metsä -menetelmää.

Satunnainen metsä -menetelmät olivat useimmin parhaiden menetelmien joukossa, joten sen viimeisin versio valitaan ensimmäisen iteraation menetelmäksi. Syötemuuttujiksi valitaan tässä alaluvussa olleita syötemuuttujia ja ne valitaan merkityksellisyyden mukaan. Syötemuuttujien rikastamisella saadaan lisää syötemuuttujia ja rikastamiseen liittyy tietämys len-

Tutkimus	Menetelmä	Syötteet
(Codina ym. 2019)	GBDT, ANN	32 kpl lentotiedoista, liikennemäärä, 9 kpl säästä
(Yin ym. 2018)	RF	lähtö- ja tulorullausjono/liikenne, tuloaika, työntö portilta
(Lee ym. 2015a)	RF, LR	portti, kiitotie, lentokoneen malli, rullausmatka, tuleva ja lähtevä liikenne, työntö portilta, lähtöaika

Taulukko 5. Rullausajan ennusteet

toyhtiön lentotoiminnasta. Seuraavassa alaluvussa kerrotaan lentoyhtiön toiminnasta tutkimuksen aihepiiriin ja syötemuuttujien rikastamiseen liittyen.

2.5 Lentoyhtiön aikataulun mukainen toiminta

Vain ilmassa olevat lentokoneet tuottavat voittoa (Wang, Wang ja Wu 2017). Tämän vuoksi lentoyhtiöt pyrkivät suunnittelemaan maassa oloajan mahdollisimman lyhyeksi. Tehokkuus toteutetaan lentokoneiden, huoltojen ja miehistöjen suunnittelulla. Lentoyhtiön toiminnan suunnittelu sisältää tyypillisesti seuraavat vaiheet (Ben Ahmed, Zeghal Mansour ja Haouari 2018) :

- Lentojen suunnittelussa päätetään, mille kentillä lennetään ja mihin aikaan, jotta saataisiin paras mahdollinen voitto.
- Lentokaluston käytön suunnittelussa päätetään, millä konetyypeillä ja -versioilla reitit kannattaa lentää.
- Reitityksen suunnittelussa päätetään lentokoneiden lentojen sarjat, jotta lennot lennetään tehokkaasti huollot huomioiden.
- Miehistösuunnittelussa kohdistetaan lentojen sarjat miehistöille, jotta heidän työvuoronsa olisivat tehokkaita ja laillisia suhteessa erilaisiin määräyksiin.

Lyhyt kääntöaika tekee aikataulussa pysymisestä haasteen. IATA:n mukaan lentokoneen kääntöaika alkaa, kun lentokone saapuu lentokentän seisontapaikalle tai portille (Oreschko ym. 2010).

Kääntöaika päättyy, kun kaikki ovet ovat suljettu, jarrut vapautettu ja lentokone on valmiina työntöön portilta (Nosedal ja Eroles 2017). Seisontapaikoilla matkustajien on noustava koneesta tai koneeseen portaita pitkin. Asematasolta matkustajat kuljetetaan lentoasemalle. Tämä aiheuttaa viivettä siirtymisissä lentoasemalta koneeseen ja koneesta lentoasemalle. (Diepen ym. 2009).

Käännön aikana tapahtuu paljon ja toimintojen on tapahduttava oikeaan aikaan. Matkustajat lähtevät koneesta, uusille matkustajille tehdään lähtöselvitys ja he menevät koneeseen. Rahti lähtee koneesta ja uusi tulee koneeseen. Matkalla kertyneet jätteet poistetaan ja uudet ruoat ja juomat tulevat koneeseen. WC tyhjennetään, vesisäiliöt täytetään, kone siivotaan ja tankataan. Mahdollinen jäänpoiston tarve tarkistetaan. Jään poisto ja esto tehdään viimeistään rullauksen aikana ennen lentoonlähtöä (Norin ym. 2011). Jotkin käännön aikaiset tehtävät voidaan tehdä samanaikaisesti, mutta osa on tehtävä tietyssä järjestyksessä (Schmidt 2017). Jotta toimenpiteet sujuisivat hyvin, sidosryhmien on tiedettävä, millä portilla ja mihin aikaan heidän pitäisi milloinkin työskennellä (Oreschko ym. 2012).

Aikatauluun sidonnaisia sidosryhmiä ovat: lähilennonjohto, lähestymislennonjohto, maaliikenteen johtaminen, porttivirkailijat, kenttäalueen kunnossapito, tulli, lentoyhtiön liikenteenohjaus ja -suunnittelu. Kaikki sidosryhmät ovat kiinnostuneita, milloin lentokone lähtee tai saapuu portille.

Suunnitelmaan aiheuttavat muutoksia lentokoneiden vikaantumiset, lentoliikenteen ruuhkat, miehistöpulat, myöhästymiset ja säätila. Nämä johtavat ongelmien ketjuuntumiseen yleensä suunnitteluketjun alempiin vaiheisiin ja aiheuttavat käännön aikaisten toimenpiteiden uudelleen suunnittelua. Aamun lennon myöhästyessä lentosarja on yleensä koko päivän ajan myöhässä (Chen ja Li 2019). Toisinaan myöhästynyt lento saadaan aikatauluun jo seuraavalla lennolla myötätuulen tai pidemmän kääntöajan avustamana. Poikkeama voi tarkoittaa, että lento on etuajassa tai myöhässä. Muutokset aiheuttavat lisäkuluja polttoaineen kulutukseen, lentokenttämaksuihin, huoltokuluihin, matkustajien majoituksiin ja ruokiin, ylityökorvauksiin ja tulonmenetyksiin uudelleen reitityksissä (Ben Ahmed, Zeghal Mansour ja Haouari 2018).

Poikkeamien havaitseminen hyvissä ajoin mahdollistaa korjausten tekemisen suunnitelmiin.

Pienemmillä myöhästymisajoilla on vähemmän seurannaisvaikutuksia muihin lentoihin, joten poikkeamat tulisi korjata mahdollisimman pian (Cook ja Tanner 2015). Jatkoyhteyksien tärkeyden vuoksi paluulennon ennustamisella Helsinkiin on suuri merkitys. Siitä huolimatta aikataulun korjaus voi olla mahdollista vain ennen lähtöä ulkokentälle. Poikkeamiin tulee reagoida etukäteen. Lentoliikenteessä on paljon poikkeamia suunnitelmista, mutta isolla tietomäärällä voidaan ennustaa hyvin keskimääräistä toteumaa ja jopa poikkeamia. Poikkeamat voivat olla säännönmukaisia, jolloin ne voidaan ennustaa. Reittilennon aikana myöhästymisaikaa voidaan pienentää lentämällä hieman suuremmalla nopeudella. Tämä aiheuttaa kuitenkin ylimääräistä polttoaineenkulutusta (Wu 2010). Korkealla lennettäessä kuitenkin koneen maksimi Mach-nopeus rajoittaa nopeuden lisäystä eikä suuria ajan säästöjä saada kuin pitkillä matkoilla. Aina lentoajan lyhentäminen ei ole kannattavaa, jos seuraavaa lentoonlähtöaikaa rajoittaakin muu liikenne SLOT-ajalla.

SLOT-toiminta perustuu EU-asetukseen ja kansainvälisiin sääntöihin. Helsinki-Vantaa on Suomessa ainoa SLOT-koordinoitu lentokenttä. Lentoyhtiön on käytettävä vähintään 80 % SLOT:istaan. Näin lentoyhtiö säilyttää SLOT:nsa seuraavallekin vuodelle. Jos lentoyhtiö ei käytä myönnetyä SLOT:ia, se jaetaan muille IATA:n ja EU:n sääntöjen mukaisesti. (Vuori ja Ahlroth 2022)

Vaikka lentokone olisi muuten valmis lähtemään portilta, ei ole sallittua lähteä aikataulua ennen, jos kaikki matkustajat eivät ole vielä saapuneet koneelle. (Eurocontrol 2020). SLOT-ajat ovat ennalta määrättyjä, mutta sään aiheuttamia poikkeamia ei määrätä ennalta. Ne ovat kuitenkin ennustettavissa. Amerikan kotimaan lennoista myöhästyi vuonna 2021 noin 18 %. Sää oli syynä noin 11–34 % myöhästymisistä (*Airline On-Time Statistics and Delay Causes* 2022). Sää vaikuttaa sekä reitillä että lähtö- ja määräkentällä. Lumisateet rajoittavat kiitoja rullaustien käyttöä sekä lisäävät lentokoneen jäätymisenestokäsittelyjä.Seuraavassa alaluvussa kerrotaan ilmailussa käytettävistä sääsanomista ja niiden käytöstä.

2.6 Säätiedot ja -ennusteet

Eurooppalaiset lentoyhtiöt noudattavat EU-OPS 1 -määräyksiä toiminnassaan. Sen liitteet D, E ja F sisältävät säähän liittyviä ja aikatauluihin vaikuttavia määräyksiä (*Official Journal of*

the European Union 2008). Olosuhteet ja sää lentokentällä vaikuttavat suoritusarvoihin ja siten käytettävään kiitotiehen ja lentomenetelmiin. Lentoonlähdöille ja laskeutumisille sallittu näkyvyys ja pilven alaraja ovat määritetty tietyin portain koneluokittain. Samoin sivutuuli- ja takatuulikomponentti on rajoitettu konetyyppikohtaisesti. Molempiin tuulikomponentteihin vaikuttavat myös kiitotien kitka-arvot ja maanpinnan epäpuhtaudet. Lentokoneen pintojen on oltava puhtaat epäpuhtauksista, kuten jäästä ja lumesta. Tarvittaessa epäpuhtaudet poistetaan ja käytetään jäänestokäsittelyä, jolla on tietty suoja-aika sään mukaisesti laskettuna.

Sää vaikuttaa monella tavalla lennon eri vaiheisiin ja lentosääpalvelusta saadaan oleelliset tiedot lennon suunnitteluun. Tarvittavien lentosääpalvelun järjestämisestä vastaa Ilmatieteen laitos ("Suomen ilmailukäsikirja" 2022). Säähavaintoja ja -ennusteita tekee Suomessa 24 lentoasemien havaintoasemaa ja ATS-yksikköä. Ilmailuun erityisesti vaikuttavat säätiedot ja ilmiöt ilmoitetaan sääsanomilla. Säätila ilmoitetaan METAR-sanomalla ja sääennuste ilmoitetaan TAF-sanomalla. Muitakin sanomia käytetään, mutta METAR ja TAF ovat pääasialliset lähteet. Lisäksi reittilennolla tarvitaan ylätuulitietoja sekä sääilmiöistä kertovia karttoja. (*Lentosääpalvelut Suomessa* 2021)

METAR-sanoma kuvaa sen hetkistä säätilannetta havaintoasemalla. Säätila vaihtelee usein sekä ajallisesti että paikallisesti. Lentosääasemat julkaisevat METAR-sanomien lisäksi 10 minuutin välein olevia säähavaintoja. Näitä säähavaintoja voi hakea säähavainnot-aineistosta, mutta niiden sisältämän tiedon luonne on yleinen eikä ne aina sovellu ilmailun tarpeisiin. METAR ja AUTO METAR julkaistaan puolen tunnin välein. Osalla asemista kaikki mittaukset ovat automaattisia. Manuaalinen säähavainto, joka on lennonjohdon yhteydessä, julkaistaan usein vain lennonjohdon aukioloaikoina. (*Lentosääpalvelut Suomessa* 2021)

Automaattinen METAR-sanoma alkaa aikaryhmän jälkeen sanalla AUTO. AUTO METAR käyttää hieman poikkeavia koodeja kuin METAR-sanoma. (*Lentosääoppia harrasteilmailijoille* 2020)

Lentosääennustaminen on luonteeltaan lyhyen ajan ennustamista, sillä pisimmät TAF-ennusteet tehdään 24 tunnin jaksolle tulevaan aikaan. Lyhyt 2–3 vuorokauden yleinen sääennuste on yleensä tarkka, mutta 5–10 vuorokauden yleisennusteet ovat epätarkempia (Karttunen ym. 2008). Tietyn hetken ennusteen lopputulos on hyvin moniosainen. Tutkimuksen kannalta kiinnosta-

vat vain tietyt osat ennustetta. Ennuste on meteorologin paras näkemys tulevasta säätilanteen kehityksestä. Tämä arvio yleensä luotettava, mutta arvio muuttuu ja tarkentuu jo seuraavassa mallin ennusteajossa, koska ennustetun hetken lähestyessä saadaan tarkempia arvioita vaikuttavista tekijöistä. (*Lentosääpalvelut Suomessa* 2021)

Sääennusteissa on aina muistettava, että sääennusteet ovat piste-ennusteita. TAF on lentopaikkakohtainen pistemäinen ennuste. Mitä pidemmälle ajalle ennustetta tehdään, sen laajemmalle alueelle se on tehtävä. Tietokoneet voivat laskee ennusteita 15 vrk eteenpäin, mutta nämä ennusteet eivät välttämättä toteudu sellaisenaan. Sääennusteen tarkkuuteen vaikuttaa myös säätilanne. Blocking-korkeapaineen aikana kesällä ennusteet ovat hyvin tarkkoja koko ennustejakson pituudelta, mutta talvella vastaava tilanne ei takaa hyvää sääennusteen tarkkuutta. Talvella inversiotilanteet ovat haastavia lämpötilaennusteille, jolloin säämalli ole kovin hyvä ennustamaan lämpötilaa. Kesällä paikalliset sääilmiöt, kuten sadekuurot ja ukkoset ovat haastavia tilanteita ja useita tunteja etukäteen ennustaminen yhteen pisteeseen on vaikeaa. Eri säämalleilla on eri osa-alueet, joissa ne ovat kohtuullisen hyviä tai huonoja. (L. Saukkonen, henkilökohtainen tiedonanto, 24.8.2021)

Choi ym. (2016) ennusti sääennusteiden avulla lentoliikenteen myöhästymistä viideksi ja yhdeksi päiväksi etukäteen sekä päivän oikeilla säätiedoilla. Vertailu eri aikajaksojen ennusteista on taulukossa 6. Pidemmillä ennusteilla tarkkuutta huonontaa sääennusteen luotettavuuden huononeminen.

Aikaikkuna	Tarkkuus
Viisi päivää aikaisemmin	26,8 %
Yksi päivä aikaisemmin	30,1 %
Oikea säätila	80,4 %

Taulukko 6. Sääennusteesta ennustaminen

TAF:n tarkoitus on ennustaa lentotoiminnalle merkittäviä sääilmiöitä ja arvoja lentopaikoille tiettyinä ajanjaksoina. TAF:n tietoja käytetään lennonsuunnitteluun. TAF:n voimassaoloajat ja muutostyypit ovat lentoliikenteen tarpeiden mukaisia. Lentopaikkaennuste sisältää seuraavat tiedot: keskituuli, vallitseva näkyvyys, merkittävät sääilmiöt, pilvisyys ja edellä mainittujen tietojen muutokset ennustusjakson aikana. TAF-sanomia laaditaan kolmen tunnin vä-

lein alkaen keskiyöllä klo 00.00. TAF:ssa ilmoitetaan samat tiedott kuin METAR-sanomissa, mutta muuttuvaa säätä ilmoitetaan tiettyjen kynnysarvojen ylittyessä. Lisäksi muutoksista ilmoitetaan muutoksen tyyppi. (*Lentosääpalvelut Suomessa* 2021)

Ennusteen syötemuuttujissa säätietojen ja -ennusteiden luokittelu tulee suhteuttaa oikein EU-OPS 1:ssä annetut ja TAF:n operatiiviset rajat huomioiden, jotta luokitus toimii oikein käytännön lentotoiminnan mukaisesti. Samoin ennusteiden aikavälit tulee käsitellä, kuten EU-OPS 1 määrää ennusteiden ajoista.

Tässä luvussa kerrottiin teoriaperusteet käytettävistä tekoälymenetelmistä ja lentotoimintaan vaikuttavista tekijöistä. Näiden tietojen ja aiempien tutkimusten perusteella hankittiin alustava tietoaineisto, jota rikastettiin tietämyksen mukaisesti. Seuraavassa luvussa kuvataan aineiston valmistelua. Luku on tärkeä osa koneoppimista, koska suuri osa ennustamisessa kuluu tiedon esikäsittelyssä.

3 Tutkimusaineiston kuvaus

Tässä luvussa kerrotaan, kuinka aineisto on hankittu, esikäsitelty ja analysoitu ennen lennon vaiheiden laskentaa. Kuten aiemmin kerrottiin, esikäsittely vie 50–75 % tutkimuksen ajasta ja on tärkeydeltään noin 20 % tutkimuksesta. Tietämyksen muodostuksen tärkeyden ollessa noin 80 %. Esikäsittelyllä pyritään saamaan syötetietojen sisältö optimiksi laskentaa varten. Vaikka tietämyksen muodostus olikin tärkeämpi vaihe, esikäsittely on välttämätön vaihe datan saamiseksi optimiksi ennusteiden laskentaa varten.

3.1 Tiedon hankinta

Tutkimuksen vaatima tieto ei ollut saatavilla yhdestä paikasta, joten tiedonhankinta aloitettiin vuosi ennen tutkimusta. Mahdollisimman paljon tietoa pyrittiin saamaan julkisista tietolähteistä, jotta tutkimuksesta tulisi julkinen.

Aikaväliksi on valittava riittävän pitkä aikajakso, jossa ovat kaikki vuodenajat. Aikaväli ei saa sisältää lähtö- tai määräasemalla kiitotien kunnostusjaksoa. Jotta liikennemäärät vastaisivat normaalitoimintaa, on koronan aika rajattu pois. Näillä perusteilla aikajaksoksi valikoitui 1.8.2018–31.7.2019. Kenttäparina käytetään Helsinkiä ja Oulua, koska kyseisellä kenttävälillä on vilkkain kotimaan lentoliikenne.

Lentoyhtiöillä ovat kaikki tarvittavat lentotiedot ja tarvittaessa tiedot ulkomaisista käyttämistään lentokentistä. Lisätietoja ulkopuolisista lähteistä tarvitaan parantamaan ennusteita. Säätiedot ja liikennemäärät ovat välttämätön lisä lentotietoihin. Lentotietoja on lisäksi saatavilla internet-sivustoillta, joissa jaetaan lentojen historiatietoja. Lentokoneissa oleva toisiotutkavastaaja lähettää tietoja lennosta lennonjohtojärjestelmään ja tuo tieto on vapaasti saatavilla. Tietoa nimitetään ADS-B:ksi. Tuo tieto olisi riittävä lentotietojen osalta, mutta vaatisi paljon esikäsittelyä ja historiatiedon saanti on yleensä maksullista. Mikäli tietoihin jää puutteita, ADS-B-tiedoilla saa kohtuullisen tarkasti lisätietoa, joka parantaa ennusteita.

Ensimmäinen tiedonhankinta alkoi Traficomilta 18.6.2020. Useita yhteydenottoja oli kahdeksan kuukauden ajan, mutta kiinnostuksesta huolimatta lentoliikenteestä vastaava viran-

omainen ei pystynyt luovuttamaan tietoja. Aiemmin tietoa on kyllä saatu, mutta nyt aika ei ollut otollinen YT- ja korona-ongelmien takia.

Finnairilta tietopyyntöön vastattiin myöntävästi puolen vuoden kuluttua tietopyynnöstä. Tämän jälkeen yhteistyö jatkui ja sujui hyvin. Tarkensimme heidän tavoitteitaan useissa kokouksissa ja datan täydennyksiä tehtiin pikaisella toimitusajalla. Kokonaisuudessaan tieto sisälsi suunnitellulta aikaväliltä sekä Oulun että Rovaniemen lennot Helsingin väliltä. Varsinainen ennuste tehtiin Oulun ja Helsingin välille, mutta Rovaniemi lisäsi tietomäärää ennustettaessa Helsingin rullausaikoja ja käytettävää kiitotietä. Lentotiedot sisälsivät 9400 lentoa, joista Oulun ja Helsingin välisiä lentoja oli 6045. Kunkin lennon tietorivi sisälsi seuraavat tiedot: päivämäärä, lennon tunnus, koneen rekisteri, konetyyppi, lähtökenttä laskukenttä, lähtöportti, aikataulun mukainen aika, portilta lähtöaika, lentoonlähtöaika, lentoonlähtökiitotie, laskukiitotie, laskuaika, portille tuloaika, aikataulun mukainen saapumisaika ja saapumisarviot. Asiakas- ja rahtimäärät oli jätetty pois datasta, jotta tietoa voidaan käsitellä julkisesti.

Finavian tietopalvelusta tulivat tiedot Helsinki–Vantaan liikennemääristä. Oikea osoite pyynnölle ratkaisi tilanteen, koska Oulun ja Helsingin lentoasemilta tietoa ei saatu viiden kuukauden kirjeenvaihdosta huolimatta. Liikennetiedot sisälsivät Helsingistä 8512 tietoriviä. Kukin rivi sisälsi tunnin välein lentoonlähtöjen ja laskujen lukumäärät. Finavian tietopalvelu täydensi tutkimuksen aikana vielä liikennemäärien tietoja Oulun ja Rovaniemen osalta, mutta tiedoista oli karsittu sotilasliikenne pois eikä yksittäisten lentojen kiitotietietoa saatu.

Ilmatieteenlaitokselta säätiedot, ohjeet ja keskustelut tulivat nopeasti ja säiden puolesta tiedonhankinta sujui valmiiksi asti muutamassa viikossa. Lisäksi vastaukset kysymyksiin tulivat nopeasti ja perusteellisesti. Helsinki–Vantaan lentokentältä on 17473 ja Oulusta 17391 METAR-sanomaa. TAF-sanomia on samalle aikavälille validointia varten Helsingistä 3386 ja Oulusta 3292 sanomaa. Lisäksi Rovaniemeltä ovat samankokoiset tietomäärät valmiina uuden kenttävälin lisäyksen testaukseen. Sanomissa on vaihtelevasti seuraavia tietoja: päivämäärä, tuulen suunta, tuulen nopeus, tuulen puuskat, ukkonen jäätäminen lumisade, sade, näkyvyys, kiitotienäkyvyys, pystynäkyvyys, pilvisyyden määrä, pilven alaraja, sumu, lämpötila, kastepiste ilmanpaine. Kolmelta kentältä oli yhteensä 68000 sääsanomaa. Kukin sääsanoma puretaan noin 55 tietosarakkeeseen

Ylätuulien historiatiedon hankkimiseen käytettiin Meteoblue-sivustoa. Tutkimukseen saatiin heiltä history$^+$ -status, jolla saa ylätuulien historian 250 hPa (FL340) ja 500 hPa (FL180) pinnoille. Tiedot ovat UTC-aikaa ja nopeusyksikkönä m/s. Ylätuulet ovat valmiiksi jaoteltuina tutkimuksen tarpeisiin tunnin välein. Mittauspaikkoina olivat Jyväskylän, Oulun ja Rovaniemen kohdat lentopinnoilta 340 ja 180. Tietoa oli yhteensä 22560 riviä.

Meteoblue luovuttu tiedot opiskelijakäyttöön korvauksetta ja he pyysivät kopion tehdystä tutkimuksesta sen valmistuttua. Lentotietojen täydennykseksi pyysin tutkimustarkoituksiin Oulun käytössä olevien kiitoteiden tiedot FlightAware-sivustolta, mutta heidän hintansa olisi ollut 3716 USD. Ilmaista METAR- ja TAF-tietoa oli saatavilla NavLost-sivustolta.

3.2 Esikäsittely ja analyysit

Ensimmäisen iteraation ennusteet lasketaan koko suunnitelmalle staattisella tiedolla. Lopullinen sovellus tehdään pääasiassa dynaamisilla tiedoilla. Syötemuuttujien analyysejä ja testaamista nopeutettiin IBM SPSS Modeler -työkalulla, joka sopii nopeaan ja automaattiseen testaamiseen. Ohjelma tekee paljon automaattista tiedon esikäsittelyä ja sillä on helppo tehdä analyysejä tiedonkäsittelyn välivaiheista. Tiedon käsittely perustuu tietovirtaan, johon lisätään valmiita toimintoja valmiilla solmu-elementeillä. Solmuja voidaan myös graafisesti muokata ja vertailla, miten muokkaus vaikuttaa lopputulokseen. Automatisointi voidaan tehdä niin pitkälle, että solmuilla valitaan paras tekoälymenetelmä, jolla lasketaan ennustemalli tiedoille ilman datan esikäsittelyä (*SPSS Modeler* 2022). Tämä nopeuttaa tiedon käsittelyä, mallin valintaa ja syötemuuttujien valintaa. Syötemuuttujien null-arvot, oikeellisuus, ääriarvot ja oikeat tietotyypit voidaan tehdä yhdellä Data Audit -solmulla.

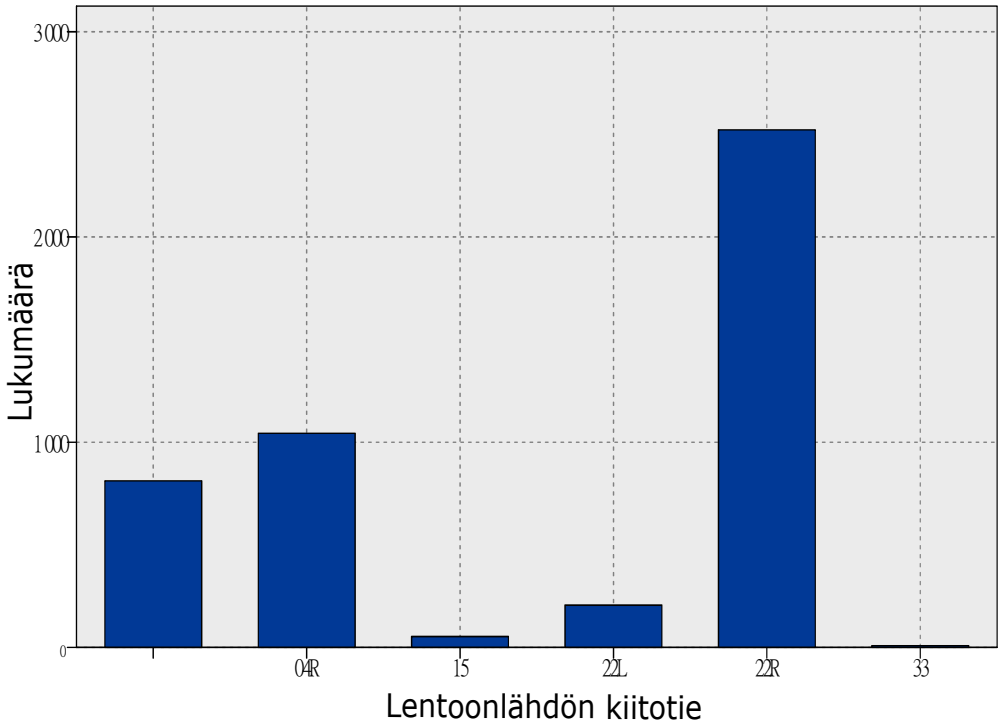
Lopullinen tarkka ennuste tehdään testaamalla ja optimoimalla parhaita ehdokkaita valituilla syötemuuttujilla ja Python SciKit-kirjaston menetelmillä. Pythonin menetelmä voidaan siirtää valmiina laskentana lopulliseen sovellukseen. Mallia rakennetaan ja arvioidaan SPSS Modelerin graafisella käyttöliittymällä ja SciKit-malli tehdään, kun on saatu parempi ymmärrys mallin toiminnasta ja syötemuuttujien vaikutuksista.

Ennen tietojen rikastamista tarkistetaan, että lähdetiedot eivät sisällä korreloivia, vääriä, poikkeavia tai tyhjiä tietoja. Jotkin poikkeamat ovat nähtävissä vasta tiedon rikastamisen

jälkeen. Tällainen on esimerkiksi lentoaika. Lentoaikaa ei ole laskettuna lentotiedoissa, vaan tiedoissa on lähtöaika sekä laskuaika. Esikäsittelyssä voidaan esimerkiksi löytää väärin merkityt ajat, rikastamisessa lasketaan lentoaika ja lentoajoista etsitään poikkeavia lentoaikoja. Poikkeamia ovat selvästi muista poikkeavat arvot. Näitä poikkeavia arvoja ei käytetä opetusdatassa, koska menetelmä laskisi kaikille lennoille keskimääräistä pienintä virhettä, jolla korjattaisiin poikkeavan arvon virhettä. Tämä vääristäisi normaaleiden lentojen ennusteita. Tällainen normaalista poikkeava tieto pitää poistaa opetusdatasta. Lennon muut vaiheet voidaan käyttää ennusteiden tekemiseen muissa kyseisen lennon vaiheissa. Poistettu vaihe rikkoo kuitenkin lentosarjan, johon poistettu vaihe kuului. Tällaista lentoa ei voida käyttää kokonaisuuden ennustamisessa. Alustavissa lentotiedoissa on 24 syötemuuttujaa ja alustavan rikastamisen jälkeen syötemuuttujia oli 64. Lopullisiin malleihin valikoitui yleensä noin 10–15 syötemuuttujaa. Rikastaminen tehdään aluksi yleisesti koko lentotiedolle ja myöhemmin rikastetaan erikseen tietoa kussakin eri lennon vaiheissa, jolloin muuttujien määrä kasvaa vielä lisää.

Ensin lentotiedoista poistettiin lennot, jotka eivät kuulu tietojoukkoon. Kahdeksan lentoa ei ollut Helsingin ja Oulun väliltä. Lennot, joiden lähtö- tai tulokenttä oli CHQ, NUE tai PMI, poistettiin. Lisäksi tiedoissa oli 24 lentoa väärällä rekisteritunnuksella (SE-MDC) tai konetyypillä (73W). Kiitotie- ja porttitiedot puuttuivat kaikista Oulun ja Rovaniemen lentokentistä eikä niitä saatu pyynnöistä huolimatta. Finavialta saatiin tieto käytetystä kiitotiestä eri tunneille eroteltuna. Helsingin kiitotietiedot olivat kattavat ja vain 812 kiitotietietoa puuttui 4645 lentoonlähdöstä ja 712 puuttui 3936 laskukiitotietiedoista. Laskeutumiskiitotie vaikuttaa lentoaikaan ja rullausaikaan. Helsingin puuttuvat kiitotietiedot jätettiin pois kyseisistä vaiheista. Yhden vaiheen poisto estää kyseisen lennon käytön kokonaislaskennasta. Kuvio 3 kuvaa lähtevien lentojen käyttämien kiitoteiden jakaumaa ja kuvio 4 kuvaa laskeutuvien lentojen käyttämien kiitoteiden jakaumaa. Suurin osa lennoista lähtee kiitotieltä 22R ja laskeutuu kiitotielle 15. Helsingin puuttuvien kiitotietietojen osuus on 18 % lennoista. Kiitoteiden käyttö ei jakaannu tasaisesti, joten tarvittaessa kiitotie-ennustetta, ennusteessa on huomioitava epätasainen jakauma.

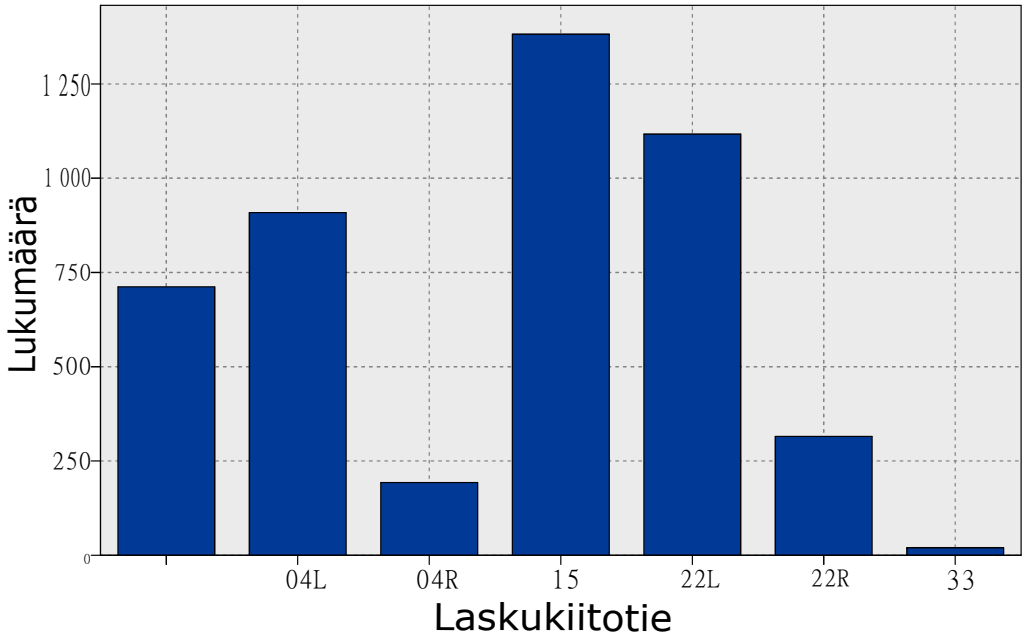
Säätiedot pitää purkaa rakenteiseksi tekstitiedosta, jotta voidaan laskea niiden sisältö. Purkaminen on tehtävä automaattisesti, johtuen sanomien suuresta määrästä. Sääsanoma on me-



Kuvio 3. Helsingistä lähtevien lentojen kiitotie

teorologin kirjoittama tekstijono. Sanomaa puretaan, kunnes kaikki tieto on käsitelty. Vain viidessä METAR-sanomassa oli kirjoitusvirheen vuoksi automaattiseen esikäsittelyyn kelvotonta tekstiä. Kyseiset virheet korjattiin käsin ennen esikäsittelyä. Purkamisessa valitaan oikean muotoista tietoa sanomasta ja onnistuneen valinnan jälkeen poistetaan tehty valinta sanomasta, jota käsitellään. Mikäli tietoa ei löydy, siirrytään seuraavan tiedon valintaan. Tuulitiedon jälkeen jatketaan käsittelyä järjestyksessä: näkyvyys, RVR, ukkonen sateet, jäätäminen, sumu, pystynäkyvyys, pilvet, lämpötila ja kastepiste. Etenemisessä on huomioitava ilmiön merkittävyys sekä niiden yhdistelmät. Esimerkkeinä luokitelluista lumisateen muodoista ovat: -SHSN, -IC, IC, SHSN, -SN, -SHRASN, SN, -SHSNRA, DRSN, SHRASN, BLSN, +DRSN, -DZSN, +BLSN, -FZDZ, -SNDZ, +SN, FZFG, -RASN, SHSG, FZDZ, RASN, -SG, -FZRA, -SNRA, SG, -FZUP, SNRA, -PL, +SNRA, SNPL, FZDZSN, PL, -FZRASN, -SHGS, FZRA, SHGS, FZUP, GS ja GR.

Rakenteiseksi purettu tieto rikastetaan seuraavassa vaiheessa. Rikastamisen tarkoituksena on

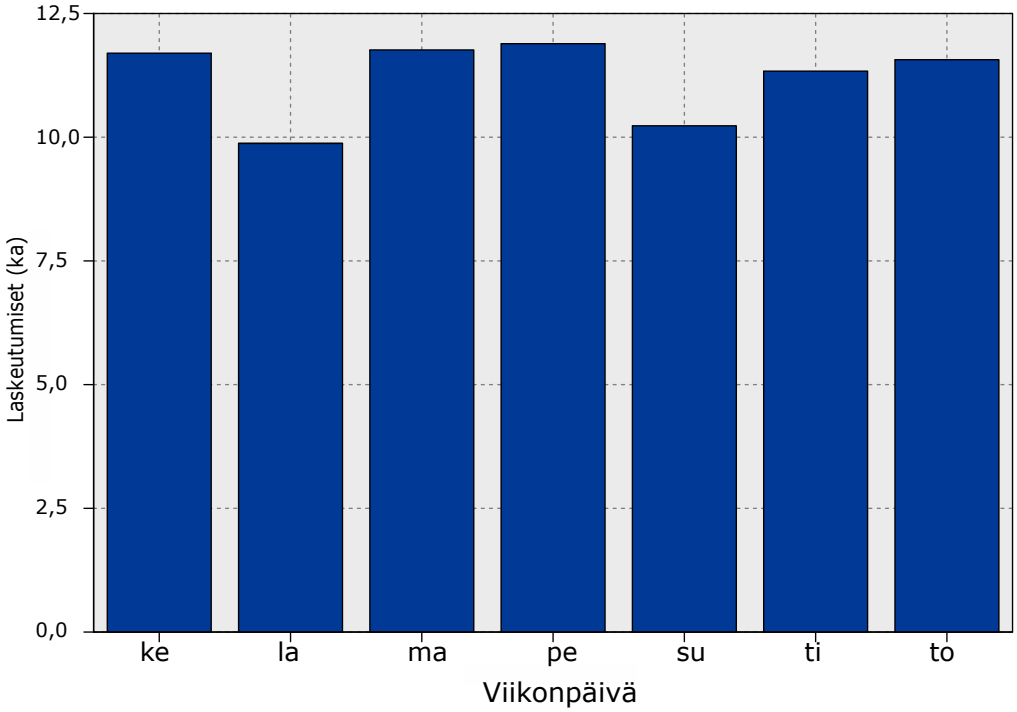


Kuvio 4. Helsinkiin laskeutuvien lentojen kiitotie

luokitella teksteistä vaihteluvälit käytännön toimintarajoihin. Samalla toteutetaan sääennusteiden mukaiset raja-arvot. Raja-arvojen mukaisesti sääilmiöt jaetaan luokkiin yhdestä viiteen. Luokiteltavat säät ovat: näkyvyys, lumisade, jäätäminen ja sivutuuli. Luokittelun raja-arvoina käytetään viranomaisen esittämiä lentotoiminnan sääminimejä näkölähestymiseen, ei-tarkkuus-lähestymiseen, tarkkuuslähestymiseen sekä huonon sään toimintaan. Luokittelu lumitiedosta on esitetty liitteessä A. TAF-sanomien eli sääennusteiden purkamisessa toistuvat samat periaatteet, mutta ajallinen vaikutus pitää jakaa jokaisesta sanomasta eri ajanjaksojen säätiedoksi. Samalla on huomioitava säämuutoksen tyyppi. Sään muutoksen tyyppi voi olla hetkellinen, jatkuva tai ajoittainen. Seuravan kolmen tunnin kuluttua julkaistavalle TAF:lle on tehtävä samanlainen jako eri ajanjaksoina tapahtuviin säämuutoksiin. Jokaisesta TAF sanomasta muodostuu näin kolmen tunnin välein tehty sääennuste tulevalle 24 tunnille. Eli yhdelle tunnille on olemassa kahdeksan eri sääennustetta, jotka ovat eri aikoina tehtyjä.

Helsingin liikennetiedoista kootaan liikennemääriä kuvaava tieto eri ajanjaksoille. Liikennemäärien vaihtelut tapahtuvat erilaisissa jaksoissa. Liikennemäärät ovat eroteltuina tulevaan ja lähtevään liikenteeseen. Liikennemäärät eivät vaihtelevat paljoa eri viikonpäivinä, kuten

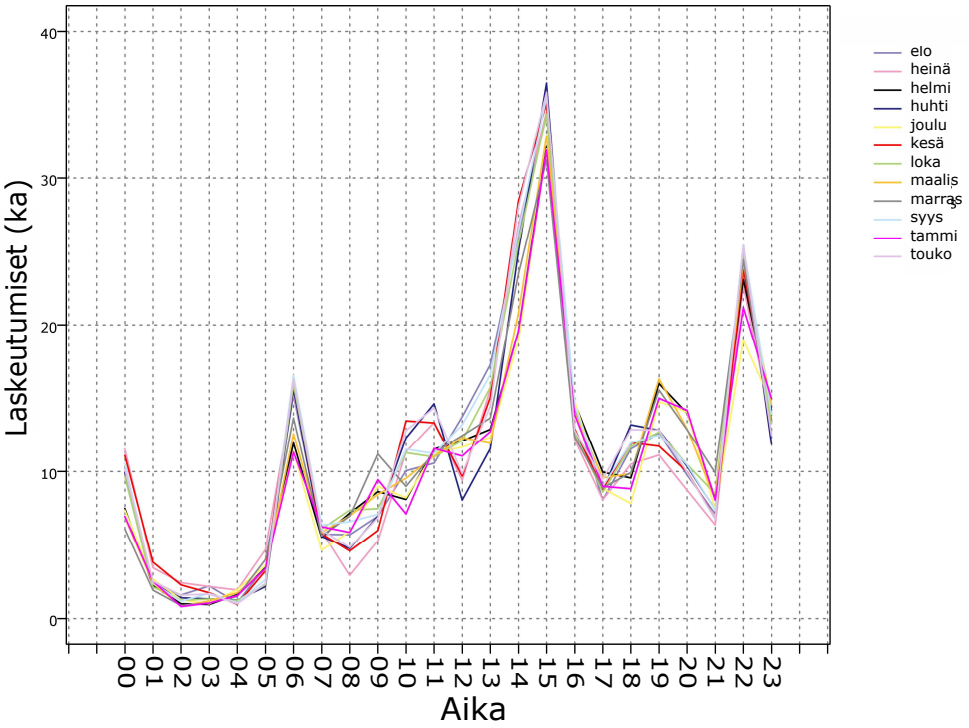
kuva 5 osoittaa. Tavoitteena on saada malli, josta saadaan luotettava tieto lennon kunkin vaiheen aikaiselle liikennemäärälle.



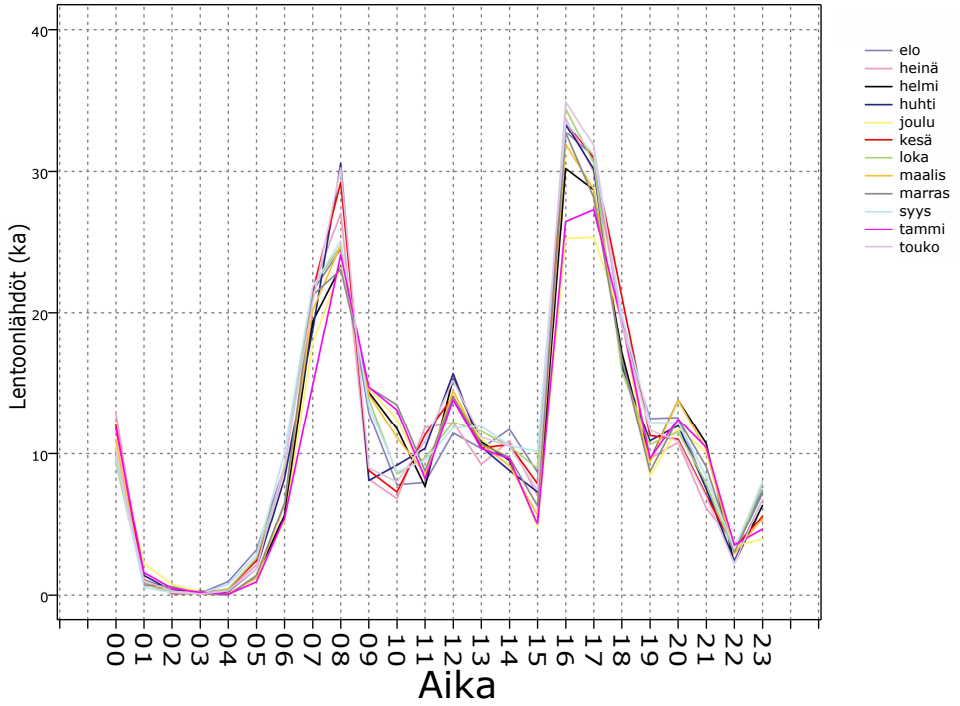
Kuvio 5. Helsinkiin laskeutuvien lentojen jakauma viikonpäiville

Seuraavaksi verrataan liikennemäärän vaihtelua vuorokauden tunteina ja kuukausina. Tyypillisesti lentoasemalla on ensin saapuvien koneiden ruuhka ja jatkomatkustajien vaihdon jälkeen tulee lähtevien koneiden ruuhka. Kuvissa 6 ja 7 ovat Helsingin tulevien ja lähtevien lentojen lukumäärät keskimäärin eri kuukausina ja kellonaikoina. Kuvista näkee tulevien koneiden ruuhkahuipun kello 15.00 (17.00 Suomen aikaa) alkaen ja matkustajien koneen vaihdon jälkeisen lähtevien koneiden ruuhkan kello 16.00 alkaen. Aamuisin on lähtevien koneiden ruuhka isompi, koska osa koneista yöpyy Helsingissä ja vastaavasti illan tulevien koneiden ruuhkan jälkeen ei enää ole lähtevien koneiden ruuhkaa. Tulevat koneet hidastavat jonkin verran maassa lähtevää liikennettä ja päinvastoin. Toisiinsa vaikuttava liikenne johtuu rajoitetusta rullauskapasiteetista sekä rullaavien koneiden kiitoteiden ylityksistä. Joskus myös lähtevä lentoliikenne aiheuttaa ilmatilassa ruuhkaa tulevalle liikenteelle ja päinvastoin.

Kuvaajista näkee, että suurin vaikuttava tekijä liikennemäärille on kellonaika. Hieman vaih-



Kuvio 6. Helsinkiin laskeutuvien lentojen jakauma eri kellonaikoina ja kuukausina



Kuvio 7. Helsingistä lähtevien lentojen jakauma eri kellonaikoina ja kuukausina

telua tulee myös kuukausittain eri kellonaikoihin, mutta ensimmäisessä iteraatiossa tutkitaan vain kellonaikoihin perustuvia liikennemääriä. Liikennemääristä tehdään kolmen muuttujan taulukko, josta haetaan muuttuja kunkin lennon liikennemäärälle kyseisen lennon vaiheen ajalle. Kolme muuttujaa ovat: lentoonlähdöt, laskeutumiset ja edelliset yhteensä. Vastaavia arvoja oli käytetty myös aiemmissa tutkimuksissa.

Ylätuulien esikäsittelyssä valitaan vaakasuorat komponentit syötemuuttujiksi. Näiden ylätuulitietojen lisäksi käytetään pintatuulesta johdettua oletustuulta. Oletustuulen muodostuksessa tehdään maanpinnan kitkan vaikutuksen poisto muuttamalla suuntaa ja nopeutta sekä tämän jälkeen oletustuulesta valitaan käytettävä tuulikomponentti.

Jotta kokonaisennustetta voidaan tutkia ja muuttaa mahdollisia virheitä tai epätarkkuuksia, lento jaetaan vaiheisiin, joilla on omat erityispiirteensä ja erilaiset merkittävät syötemuuttujat. Näitä erityispiirteitä tukemaan on lisätty rikastettua tietoa vaiheiden laskennan yhteydessä. Tämä rikastaminen kuuluu esikäsittelyyn, mutta se tapahtuu suunnittelutieteen menetelmän mukaisesti iteroinnin yhteydessä, kun tietämys lisääntyy käsiteltäessä vaihetta tietopohjasilmukassa. Tällainen palaaminen edelliseen vaiheeseen parantaa menetelmän mukaisesti tietoa sekä tietämystä.

Seuraavassa luvussa tehdään ennustemalli kustakin lennon vaiheesta. Laskenta on ensimmäinen iteraatio, jossa todetaan menetelmän toteuttamiskelpoisuus. Vaiheiden ensimmäinen summaaminen kertoo toteutuksen jatkokelpoisuudesta. Suurimmat muutokset voidaan tehdä ensimmäisellä iteraatiolla, mutta toinen iteraatio mahdollistaa ennusteen tarkentamisen ja yleistyvyyden parantamisen. Kaikki tämä tehdään opetusdatalla ja testidata pidetään kaikilta iteraatioilta sivussa lopullisen menetelmän vahvistamiseksi.

4 Vaiheiden opetus

Yhden lennon kokonaisaika ennustetaan neljässä eri vaiheessa. Vaiheet ovat kääntöaika, rullausaika kiitotielle, reittilento ja rullaus tuloaseman portille. Lähtöajasta voidaan laskea vaiheittain koneelle ennustettu aikataulu. Lentokoneen koko päivän lennot voidaan ennustaa kokoamalla näitä neljän vaiheen sarjoja. Syötemuuttujiksi valitaan eri lähteiden mukaan hyviksi koettuja syötteitä. Ennustemallin opetuksessa saatavaa tietämystä hyödynnetään uusien syötemuuttujien rikastamiseen.

Tämä rikastaminen voidaan luokitella jo tietämyksen muodostuksen toiseen iteraatioon, vaikka muuten tutkimus etenee tässä vaiheessa ensimmäisen iteraation tiedonlouhinnassa. Kolmen silmukan mallissa tietopohja-silmukan käsittely aiheuttaa merkityksellisyys-silmukan kautta tietämyksen lisääntymistä. Tästä johtuen ennusteen opetuksen yhteydessä tutkitaan myös permutaation tärkeyden kautta piirteiden merkityksiä.

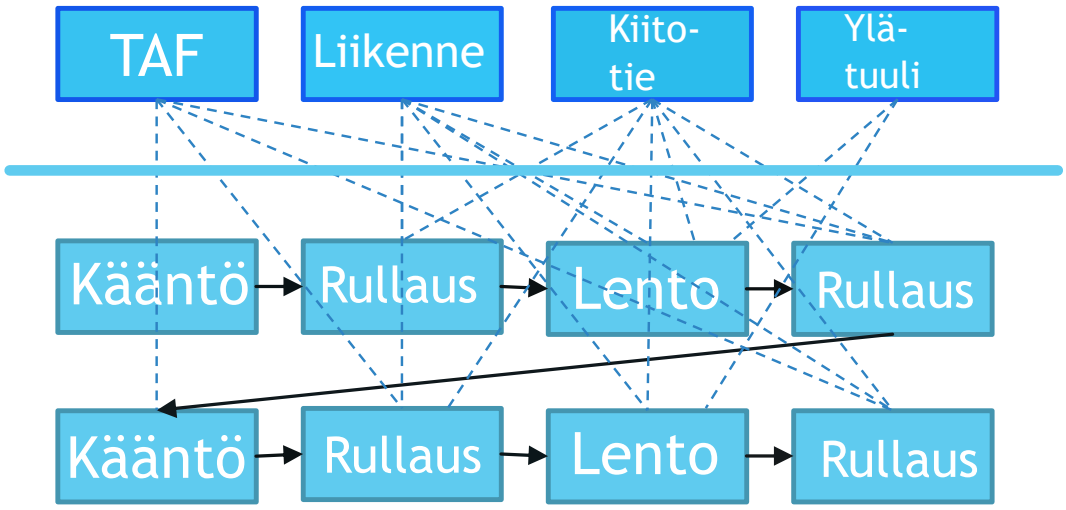
Käytettävien syötemuttujien on oltava mahdollisimman yksinkertaisia kerätä ja mahdollisimman yhtenäisiä eri vaiheiden kesken. Jakamalla kokonaisuus osiin voidaan seurata, mikä osaennuste vaatii lisää opetusta, jotta kokonaisuus saadaan toimimaan halutulla tavalla. Tällaista vaiheisiin jaettua ennustamista käytettiin aiemmin, kun tekoälylle opetettiin lentämään lentokonetta poikkeavissa olosuhteissa (Baomar ja Bentley 2017).

Lähtö- ja tuloaikojen väliseen aikaan vaikuttavat monet tekijät. Eri vaiheissa käytetään osin samoja syötemuuttujia ja osin erityisesti kyseiseen vaiheeseen sopivia syötemuuttujia. Vaiheiden myöhästymisen ennustaminen ei suoranaisesti ole tutkimuksen tavoitteena, vaan lentokoneen aikataulun ennustaminen. Tutkimuksessa ennustetaan tuleva aikataulu ja siitä voidaan arvioida, tuleeko myöhästymisiä ja onko niillä merkitystä kokonaisuuteen.

Ensimmäinen vaihe on portilta lähtöpaikalle eli kiitotielle kuuluva osa. Toiseen vaiheeseen kuuluu lentoaika, joka sisältää lentoonlähdön jälkeisen lähtöreitin, reittilennon ja määräkentän tuloreitin. Tuloreitti päättyy laskuun tuloaseman kiitotielle. Kolmanteen osioon kuuluu väli kiitotieltä lentoaseman tuloportille. Näihin kolmeen vaiheeseen vaikuttavat käytössä oleva kiitotien lentoonlähtöä ja laskua varten. Neljäntenä vaiheena ennusteeseen lisätään koneen kääntöaika. Aamun ensimmäisellä lennolla kääntöaika on koko edellisen yön aika ennen ko-

neen ensimmäistä lähtöä. Käytännössä se tarkoittaa aamun valmiutta lähteä aikataulun mukaisesti. Kun ennusteet voidaan tehdä lentokentältä toiselle ja lisätään neljäntenä vaiheena ennuste kääntöajasta, saadaan tieto meno-paluulentojen ajasta portilta takaisin portille. Ennuste kääntöajasta sisältää ennusteen mahdollisista myöhästymisistä lähtöajasta. Aamun ensimmäistä lento edeltää yön yli kestävä kääntöaika, jota kutsutaan aamun lähtövalmiudeksi.

Poikkeuksellista tässä kokonaisennusteessa on, että tutkimuksessa kehitettävä ennustemalli käyttää syötemuuttujina muita ulkopuolisia ennusteita. Nämä ulkopuoliset syöte-ennusteet ovat pääasiassa muiden tekemiä ennusteita. Jotkin ulkopuoliset syöte-ennusteet toimivat syötemuuttujina kaikille lennon vaiheille. Tutkimuksessa kehitettävä ennuste tulee vaikuttamaan vaihe vaiheelta seuraavaan ennusteeseen. Ennusteiden suhteet ovat kuvassa 8. Siinä viivan yläpuolella olevat sää-, liikenne-, kiitotie- ja ylätuuliennusteet ovat ulkopuolisia syöte-ennusteita, jotka ovat syötemuuttujina useille lennon vaiheille. Viivan alla olevat vaiheiden ennusteet tapahtuvat peräkkäin ja vaikuttavat aina seuraavaan vaiheeseen. Esimerkiksi kääntöajan ennuste vaikuttaa rullausaikaan kiitotielle, josta lähtöaika vaikuttaa lentoaikaan. Kiitotie-ennuste vaikuttaa portilta kiitotielle rullaukseen, reittilennon lähtöreittiin ja tuloreittiin sekä laskun jälkeiseen rullaukseen.



Kuvio 8. Ulkopuoliset syöte-ennusteet ja vaiheiden ennusteet suhteineen

Jotta vaiheiden ennusteet saisivat mahdollisimman oikean tiedon mallin opetukseen, ulkoisten syöte-ennusteiden tilalla käytetään opetuksessa todellisia arvoja. Lopullisessa sovelluksessa ulkoisten syöte-ennusteiden tarkkuus tulee vaikuttamaan opetetun mallin ennusteen

tarkkuuteen. TAF-ennusteet ovat meteorologin tekemiä ja niiden tarkkuus paranee ennusteajan lyhentyessä. Toinen ulkoinen syöte-ennuste, johon ei voi vaikuttaa, ovat ylätuulet. Ylätuulien ennuste on laskettavissa melko tarkasti. Liikenne-ennuste on jo sääntöperusteisena kohtuullisen tarkka ja sitä voidaan tarkentaa tekemällä ennustemalli, jossa on kellonaika, viikonpäivä ja kuukausi syötemuuttujina. Kiitotien ennustaminen on tärkein tekijä, jotta nuo neljä lennon vaihetta voidaan luotettavasti ennustaa. Näiden ulkoisten syöte-ennusteiden arviointi ja käyttö tulee tarkentumaan seuraavan iteraation tavoitteen mukaisesti.

Syötemuuttujien valinnan yhteydessä on tutkittava, miten valitut lähdetiedot korreloivat keskenään. Kahden voimakkaasti korreloivan syötemuuttujan käyttöä tulee välttää, koska ne edustavat samanlaista muuttujaa. Sellaiset syötteet, joilla on pieni merkitys ennusteelle, voidaan poistaa. Kuitenkin poistossa on muistettava, että kyseessä voi olla kahden korreloivan syötemuuttujan harha. Syötteiden vähentäminen poistaa turhaa moniulotteisuutta mallista. Tämä pienentää tarvittavan datan kokoa ja parantaa muiden syötemuuttujien vaikuttavuutta. Laskennan ohessa tehdään syötemuuttujien valintaa ja saadaan lisää tietämystä tiedonlouhintaa varten. Näin suunnittelutieteen silmukat toimivat vaikuttaen kokonaisuuteen, vaikka tässä luvussa käsitelläänkin ennusteen optimointia ja mallinnusta. Samalla, kun edetään vaihe kerrallaan ennusteiden laskemisessa, saadaan lisää tietämystä syötemuuttujien rikastamiseen. Syötemuuttujien merkitystä ei voida todentaa ennen kuin on testattu ennustemallin opetusta, koska ennustemallin permutointi antaa tiedon kunkin syötemuuttujan merkityksellisyydestä. Sikäli tietämyksen muodostuksesta alkaa jo seuraava iteraatio, kun ennustetta vielä mallinnetaan aiemmassa iteraatiossa.

Seuraavissa alaluvuissa opetetaan kunkin vaiheen ennuste, kunnes kyseinen vaihe täyttää tavoitellun tarkkuusvaatimuksen. Myöhemmin vaiheiden ennusteet yhdistetään kokonaisennusteeksi. Vaiheista poiketen aluksi tutustutaan käytössä olevan kiitotien ennustamiseen.

4.1 Käytössä oleva kiitotie

Kiitotietä ei ennusteta ennen, kuin on opetettu alemman tason ennusteet tarkoilla syötetiedoilla. Alustava arvio kiitotie-ennusteesta voidaan kuitenkin tehdä. Tässä vaiheessa testattiin lentoonlähtökiitotien ennustetta. Ennuste tulee olemaan muista poiketen luokittelija. IBM

SPSS Modeler suosituksen mukaisesti testattiin eri menetelmiä ja saatiin seuraavat tarkkuudet:

- RandomForestClassifier (tarkkuus 1,0),
- XGBClassifier (tarkkuus 1,0)
- GradientBoostingClassifier (tarkkuus 0,87)
- KNeighborsClassifier (tarkkuus 1,0)
- MLPClassifier (tarkkuus 0,98)

XGClassifier-luokittelija on luonnollinen alustava valinta. Kiitotietieto puuttuu 17,5 %:sta lentoonlähtöjä, joten lähdetietoja lisättiin käyttämällä myös Rovaniemelle lähteviä lentoja. Yleistyvyyttä pyritään parantamaan 10-ositetulla ristiinvalidoinnilla RandomizedSearchCV:tä käyttäen. Luokittelijan käyttämät hyperparametrit ovat liitteessä B.

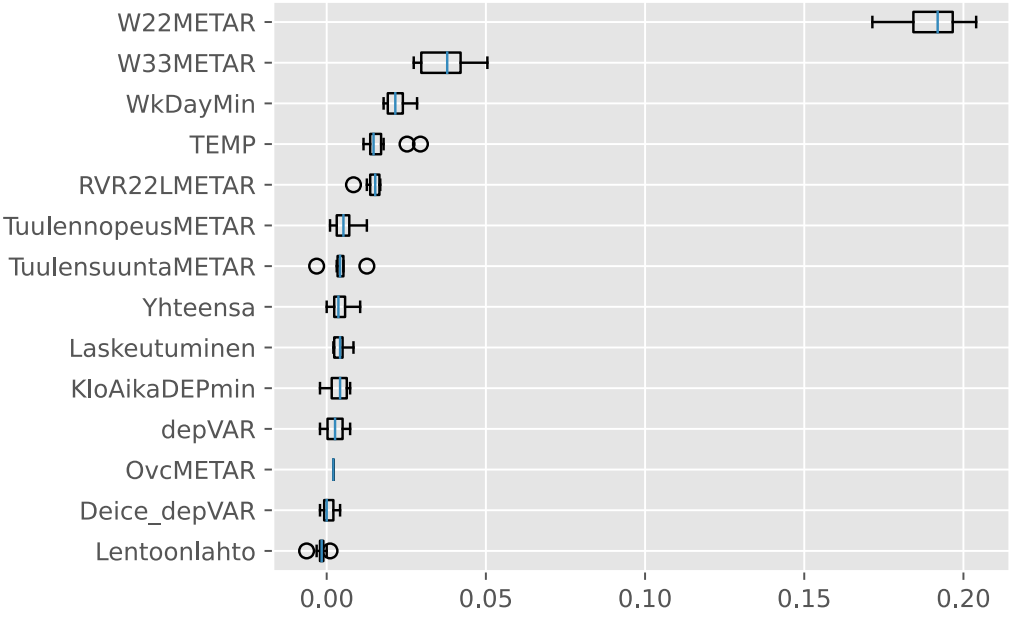
Helsingin lentokentällä on melunvaimennusmenetelmät, jotka sääntelevät kiitoteiden käyttöä. Niissä määrätään ensisijaiset kiitotiet sää- ja kiitotieolosuhteiden mukaan (AIP AD 2.21.1). Ouluun on myös ensisijaisen kiitotien määräykset kellonajan mukaan. Näissä määräyksissä käytettävä tieto tulee sisällyttää syötemuuttujiin.

Opetusdatan ennusteet täsmäsivät täysin vastedataan, joten on olemassa riski ylioppimisesta. Tilanteen monimutkaisuuden vuoksi olisi syytä vielä lisätä lentojen lukumäärää ja käyttää syväoppimisen menetelmää. Lentoja voidaan lisätä lentoyhtiön tiedoista moninkertaisesti kiitotien luokitteluun ja Helsingin rullausaikojen ennusteisiin, käyttäen pohjoiseen suuntautuvia lentoja. Muiden suuntien lennoille tulisi lisätä muuttuja, joka kertoo kohteen suunnan suhteessa kiitotien suuntaan. Tutkimuksen aluksi pitää keskittyä lennon vaiheiden ennusteisiin, jotka ovat perusta kokonaisuudelle ja ratkaisevat onko kokonaisuutta mahdollista rakentaa.

Syötemuuttujina käytetään vahvojen korrelaatioiden ja merkityksettömien muuttujien poiston jälkeen kuvan 9 syötemuuttujia.

Kiitotien 22 tuulikomponentin merkitys on 0,19. Se on selkeästi merkittävin tekijä, kun kiitotien 33 tuulikomponentin merkitys on 0,04. Myös viikon hetkellä eli tiettyinä aikoina tiettynä viikonpäivinä on merkitystä, mikä saattaa kuvastaa melu- tai ruuhka-ajan menetelmiä.

Lämpötila itsessään ei liene olennainen kiitotievalinnassa, mutta se voi kuvastaa muita merkittäviä tekijöitä. Yhdestä tuulitiedosta on rikastamalla saatu neljä tuulisyötettä, jotka kaikki olivat mukana ennusteessa. Voisi olla parasta tyytyä tuulen osalta vain yhteen tuulitietoon, joilla on suurin merkitys ennusteessa. Näkyvyyden muodot liittyvät käytössä oleviin lähestymismenetelmiin, jotka taas liittyjät eri kiitoteiden laitteisiin ja minimeihin. Jotta ennuste voitaisiin tehdä sääennusteesta, tulisi huonon näkyvyyden luokitus vastata lähimpiä lentotoiminnan raja-arvoja.



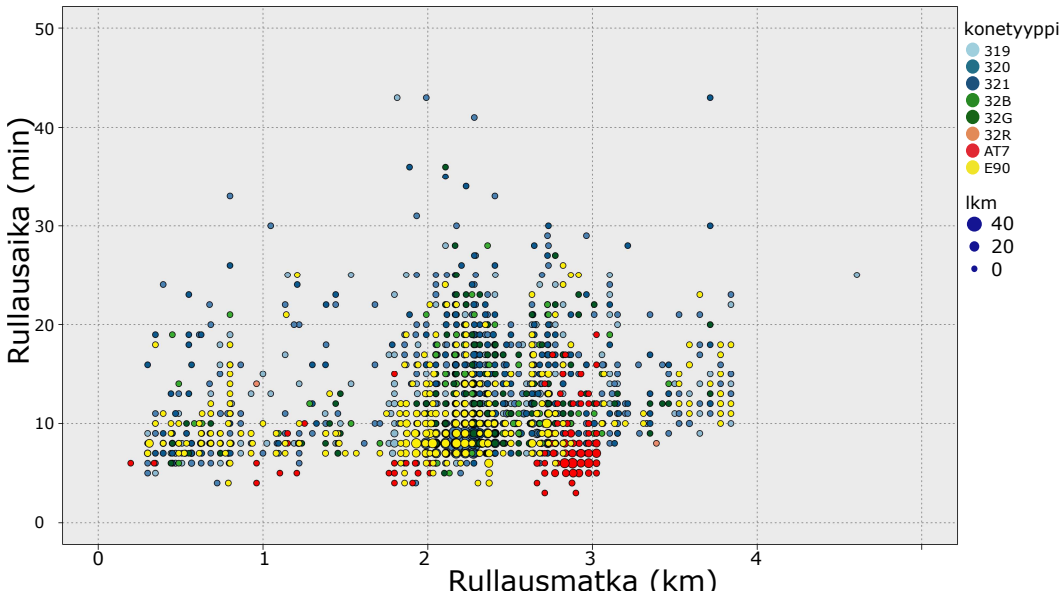
Kuvio 9. Kiitotie-ennusteen syötemuuttujien merkittävyydet

Seuraavissa alaluvuissa ennustetaan vaihe vaiheelta meno-paluulento Helsinki–Oulu välillä. Vaiheet käsitellään toteutumisjärjestyksessä alkaen Helsingin lähtöportilta. Jotta voitaisiin vertailla saman vaiheen tapahtumia Oulusta lähtien, käsitellään Oulun vastaava vaihe aina Helsingin vaiheen jälkeen. Kukin vaihe sisältää syötemuuttujien sekä ennusteen arvioinnin. Seuraavassa alaluvussa käsittely alkaa rullausajan ennusteella Helsingistä ja Oulusta.

4.2 Rullausaika lähtöportilta lentoonlähtöpaikalle

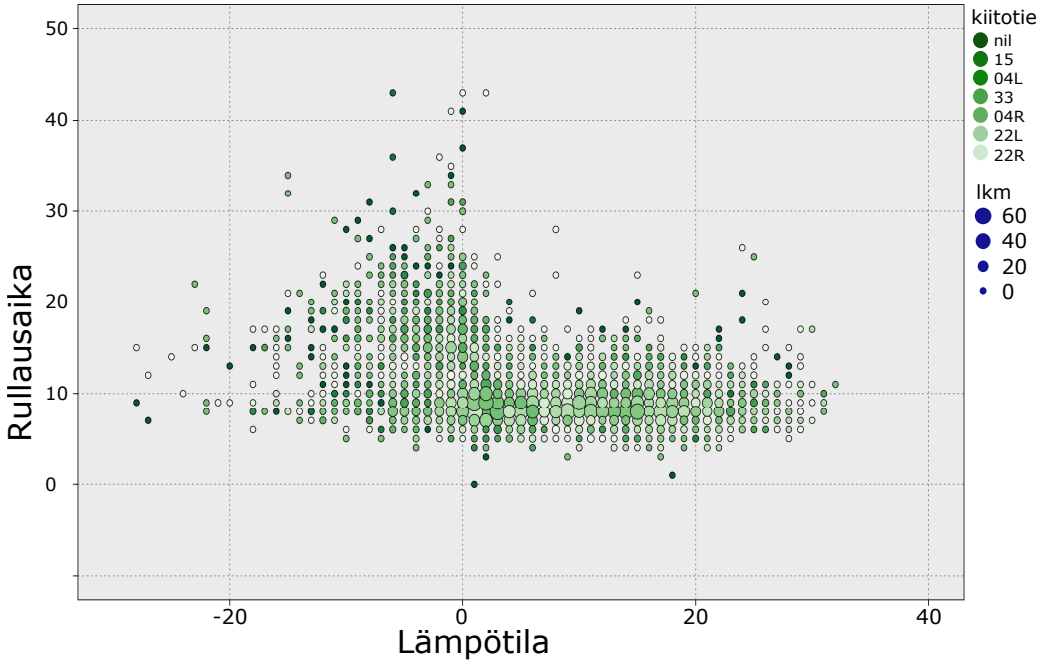
Ensimmäisen vaiheen eli rullausajan ennustetta käytetään lentoonlähtöajan laskentaan. Lähtökiitotie ja lähtöportti määrittävät, minkä tyyppinen rullaus on portilta kiitotielle. Vaikutta-

via tekijöitä ovat matka, käännökset, lähtöpaikan tyyppi ja konetyypille ominainen lentoonlähtöpaikka. Kaikkia portteja ja seisontapaikkoja ei enää ole käytössä ja uusia on tullut tilalle. Tutkimuksessa käytetään nykyisiä portteja ja seisontapaikkoja. Vanhat portit datasta on muutettu lähimpään uutta vastaavaan porttiin. Rullausmatkojen pituudet ovat mitattu jokaiselta portti- ja seisontapaikalta jokaiselle kiitotielle. Matkan laskentaan käytetään yleisesti käytettyjä rullausreittejä ja konetyyppikohtaisesti on käytetty lyhennetyn kiitotien lähtöpaikkaa. Erilaisia yhdistelmiä on 171 erilaista pituutta, joten sääntöpohjaista ratkaisua tarvitaan matkojen laskennassa. Rullausmatkan pituus vaihtelee välillä 0,2–4,6 km ja keskimääräinen matka on 2,0 km. Rullausmatkan suhdetta rullausaikaan kuvataan kuvaajassa 10. Matkan kaksinkertaistuessa rullausaika ei kaksinkertaistu. AT7-konetyypillä matka on usein pitkä, mutta rullausaika ei välttämättä kasva. Pelkällä rullausmatkalla ei siten voida ennustaa rullausaikaa.



Kuvio 10. Rullausajan ja rullausmatkan suhde

Toinen mielenkiintoinen kuvaaja on lämpötilan ja rullausajan suhde. Lämpötilan laskiessa pakkasen puolelle rullausaikojen hajonta kasvaa jo ennen pakkasen rajaa. Syynä voi olla lumi, kitka tai jäänpoisto. Kuvasta 11 näkee, että rullausajat ovat jakautuneet tasaisesti eri lämpötiloille, jos on lämpöisempää kuin 3 astetta. Lähellä nolla lämpötilaa alkaa ilmaantua enemmän pitkiä rullausaikoja. Olosuhteet vaikuttavat rullausnopeuteen, mutta kitkatiedot ovat jätetty pois, koska oletuksena on, että liikennealue on aina hyvässä kunnossa.

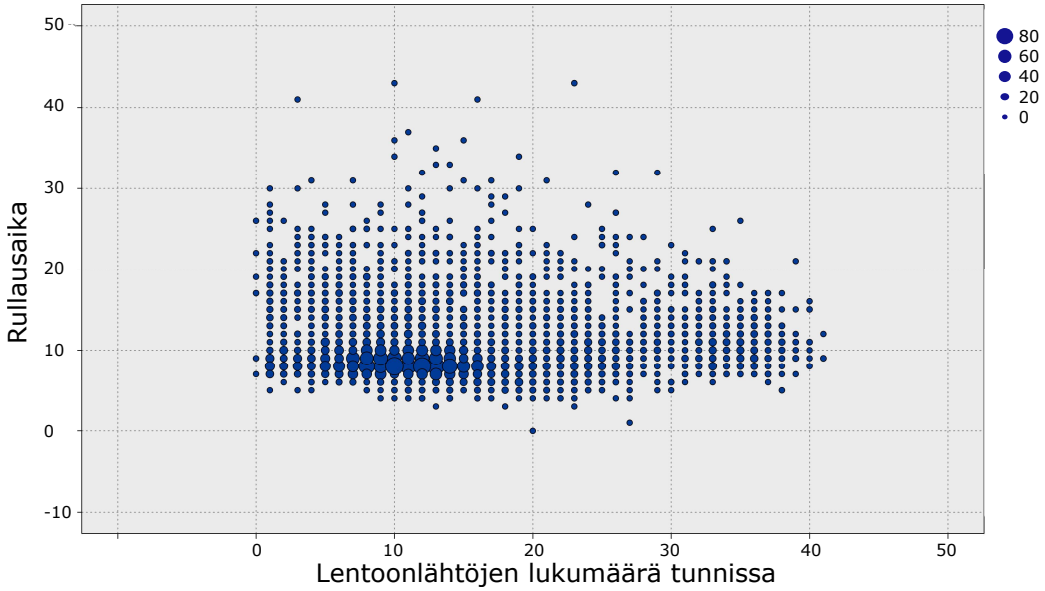


Kuvio 11. Rullausajan ja lämpötilan suhde

Liikenteen vaikutus ja ruuhka-ajat voisivat olla luonnollisia rullausaikaa pidentäviä tekijöitä. Kuviossa 12 ei kuitenkaan suurilla lentoonlähdön liikennemäärillä ole vaikutusta rullausaikojen kasvuun. Kiitotielläkään ei näytä olevan merkittävää vaikutusta rullausaikaan. Liikenteelle ei tule odotusaikoja ruuhka-aikana eli rullauksen aloitusta säädellään hyvin lennonjohdosta, jottei tule turhia odotusaikoja maassa.

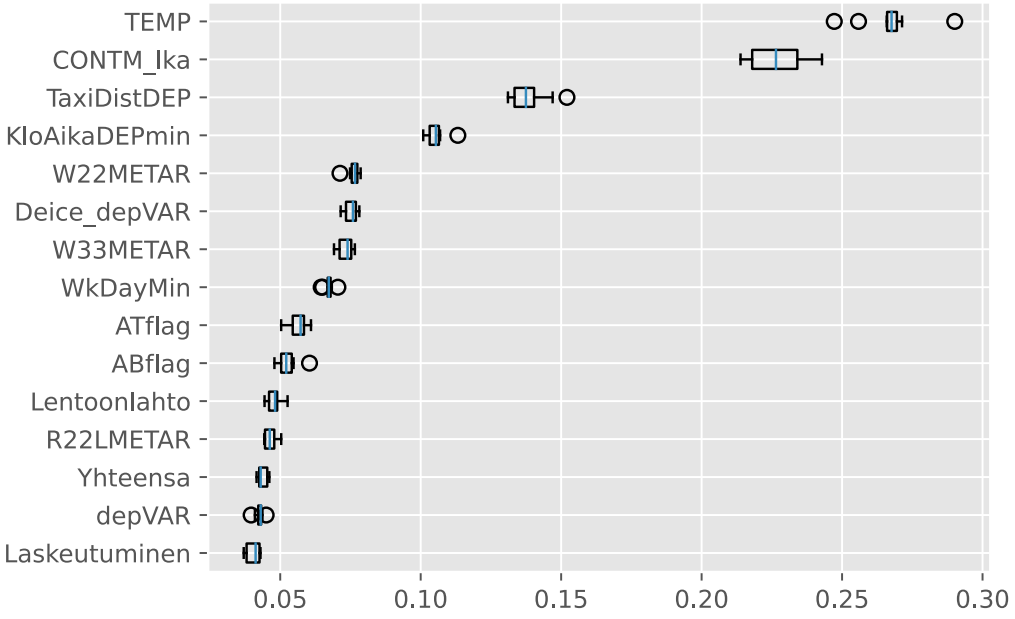
Rullausaikaan vaikuttavat monet tekijät. Ensimmäisen iteraation syötemuuttujien merkittävyydet ovat kuvassa 13. Helsingissä lämpötila näyttää vaikuttavan 0,27 verran tarkkuuteen, mutta kuten aiemmin todettiin, lämpötila voi olla jonkin muun muuttujan kanssa vahvasti korreloiva muuttuja. Lumen ja jään esiintymien vaikutus on luonnollinen vahva syy rullausaikaan (0,23). Tarvittaessa CONT-muuttujan luokittelurajoja muuttamalla seuraavassa iteraatiossa voisi saada parempia vaikutuksia ennusteeseen. Rullausmatkan vaikutus on 0,14, mutta matkan vaikutukseen voisi lisätä muita tekijöitä, koska matkan vaikutus ei ollut lineaarinen ja sen vaikutus vaihtelee konetyypeittäin.

Lopuksi poistetaan ääriarvoja eli rullausajoista on poistettu Helsingin yli 30 minuutin ja Oulun yli 15 minuutin rullaukset, jotka poikkeavat selvästi tavanomaisista rullausajoista. Muutoin nuo poikkeavat ajat vääristävät normaalin toiminnan tapahtumia pakottamalla ennus-



Kuvio 12. Rullausajan ja liikennemäärän suhde

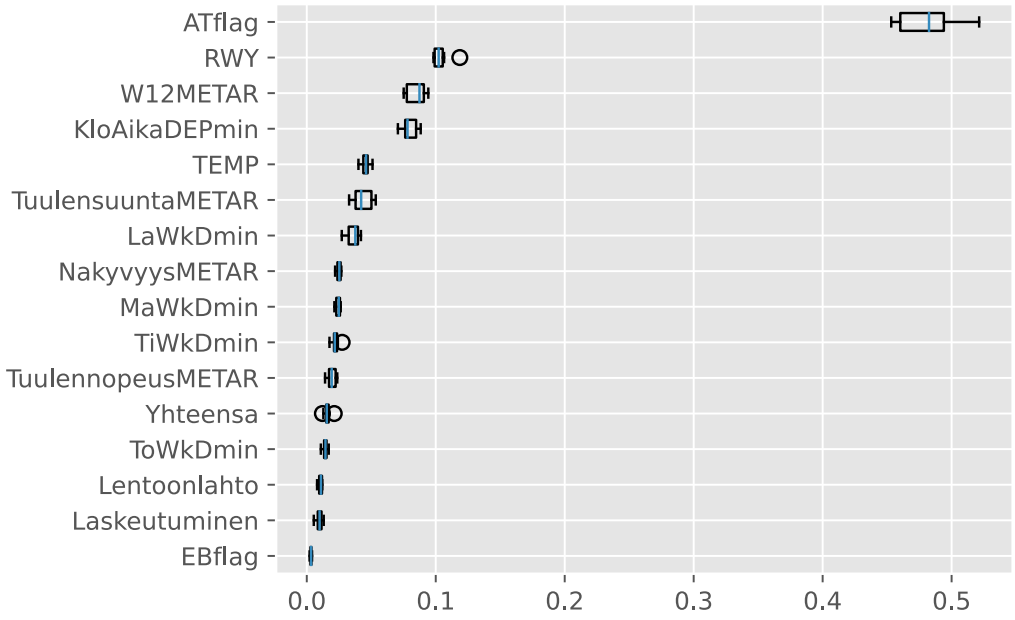
tetta huomioimaan poikkeavien tapahtumien aikoja kokonaisvirheen laskentaan. Kellonaika ennustaa ruuhkan määrää. Säätiedot ennustavat rullausteiden ja kiitotien kuntoa sekä jäänestokäsittelyn tarvetta ennen lentoonlähtöä.



Kuvio 13. Helsingistä lähtevän liikenteen rullausajan syötemuuttujien merkittävyydet

Lentoonlähdön kiitotie ja porttitiedot puuttuvat Oulun osalta, mikä on merkittävä puute. Ou-

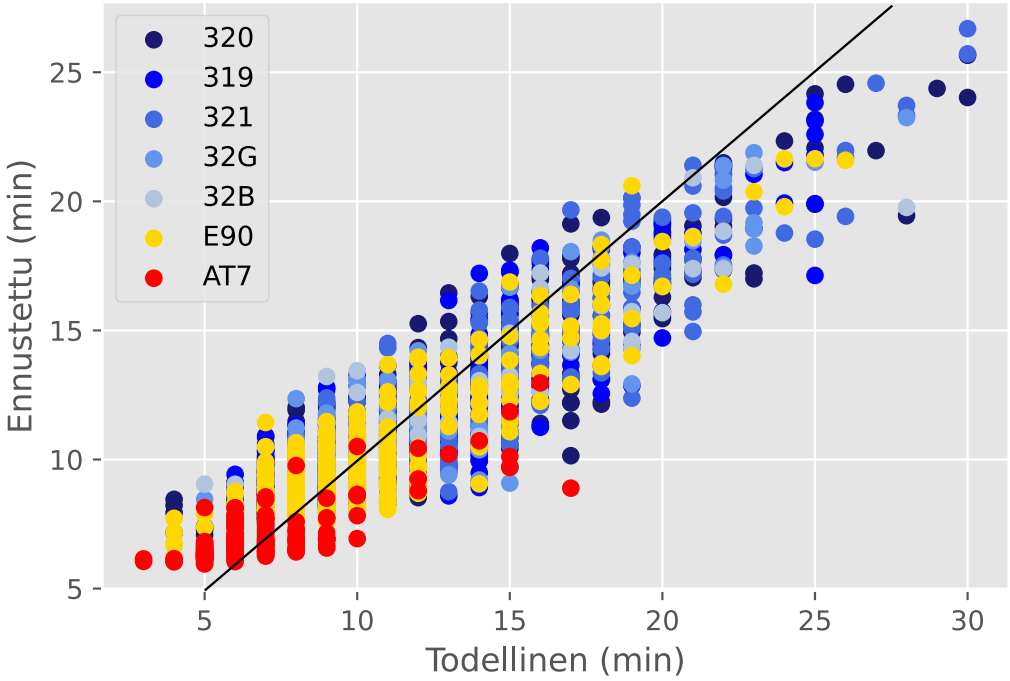
lussa ei ole tietoa käytössä olevasta kiitotiestä, lähtöportista eikä liikennemääristä, joten rullausmatkojen ja ruuhkan arvioiminen on jätetty Oulussa pois. Kiitotie-syötteen puuttuminen korostaa tuulimuuttujan merkitystä. Viikonpäivillä on enemmän merkitystä (0,02–0,01) kuin Helsingissä. Näkyvyyden merkitys 0,02 on Oulussa korostunut, koska Oulussa on vain yksi kiitotie, jonka toinen pää sallii huonon näkyvyyden lähestymiset. Helsingin erilliset kiitotiet mahdollistavat lentoonlähdöt, vaikka samaan aikaan olisi lähestyviä lentokoneita loppulähestymisvaiheessa. Kuusi syötettä oli säähän liittyvää. Oulussa tehdään jään poisto- ja estokäsittely lähtöportilla tai seisontapaikalla, joten olosuhteilla ei ole siellä isoa merkitystä. Oulun syötteiden merkittävyydet ovat kuvassa 14.



Kuvio 14. Oulusta lähtevän liikenteen rullausajan syötemuuttujien merkittävyydet

Helsingin rullausajat ovat 10,5 minuuttiaja Oulussa 6,5 minuuttia. Vastaavasti keskimääräiset virheet opetusdatalla ovat Helsingissä on 1 minuuttia 22 sekuntia ja Oulussa 48 sekuntia. Pearsonin korrelaatio on Helsingissä opetusdatalla 0,911 ja Oulussa 0,753. Oulun ennuste vaatisi lisää analysointia parempien syötemuuttujien rikastamiseksi. Kyseistä korrelaatiota voidaan kuitenkin vielä pitää hyvänä. Pidempi rullaus aika näyttäisi antavan isomman absoluuttisen virheen, mutta lyhyemmässä ajassa tai Oulun ennusteessa on enemmän hajontaa. Virheen suuruudessa tulee huomioida, että todelliset rullausajat ovat ilmoitettu kokonaisina minuutteina eivätkä sekunnin tarkkuudella. Kuvio 15 kuvaa rullausaikoja Helsingissä. En-

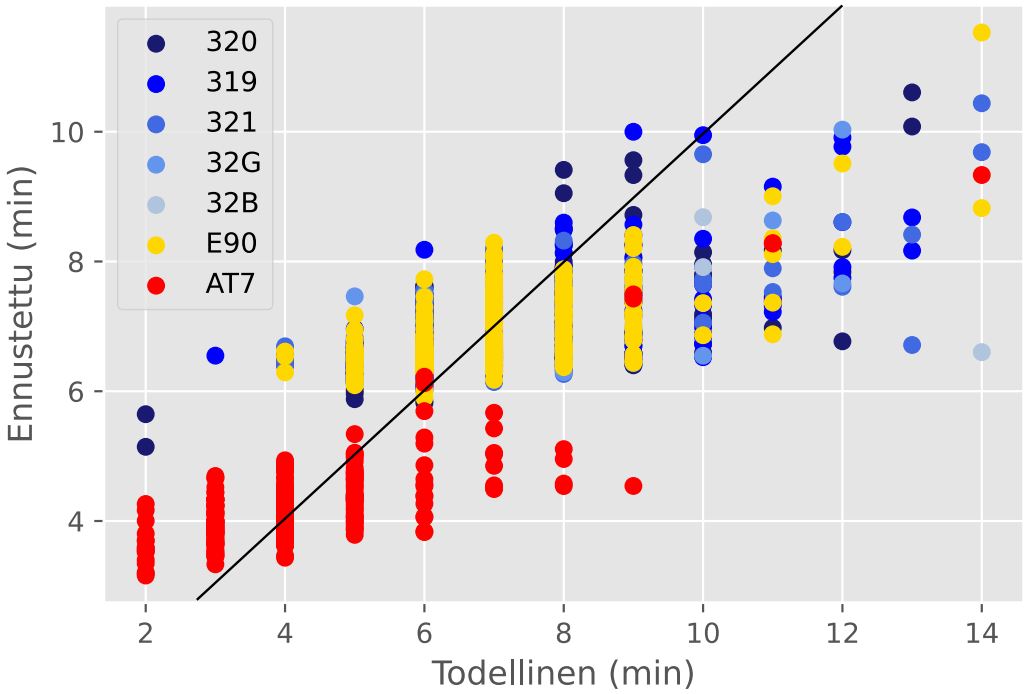
nusteet Helsingissä ovat opetusdatalla erittäin hyviä suhteessa monimutkaiseen tilanteeseen. Konetyypeittäin voidaan havaita eroa pitkien rullausaikojen ennusteissa. Pitkät rullausajat ennustetaan yleensä liian lyhyiksi. Ilmeisesti yksittäiset poikkeamat eivät ole täysin ennustettavissa ja niiden ennustaminen vaatisi syväoppimisen menetelmän testaamista.



Kuvio 15. Helsingistä lähtevän liikenteen rullausaikojen ennusteet (N=2665), $\rho_{X,Y}$= 0,911

Kuvio 16 kuvaa rullausaikoja Oulussa. Jälleen huomioida, että todelliset lennon vaiheiden ajat ilmoitetaan aina kokonaisina minuutteina eikä sekunnin tarkkuudella. Tällöin toteutuneet ajat sisältävät keskimäärin 15 sekunnin virheen oikeaan aikaan verrattuna, mikä huonontaa korrelaatiota. Oulun todellisten rullausaikojen epätarkkuus suhteessa rullausajan pituuteen aiheuttaa ison suhteellisen virheen ja vaikeuttaa ennusteen tarkkaa tekemistä. Toisaalta lyhyillä rullausajoilla ei tule isoja virheitä kokonaisuuteen.

Oulun portilta kiitotielle ennuste on toisen tyyppinen kuin Helsingin ennuste. Oulussa rullausreitit ovat lyhyempiä ja muuta liikennettä on vähemmän. Ennustetta huonontaa se, ettei Oulusta ole kiitotietietoja. Helsinkiin verrattuna aineisto on pienempi (2203 kpl), koska Helsingistä lähtevät lennot sisältävät rullaavat koneet Ouluun sekä Rovaniemelle. Oulussa kiitotien syötemuuttujana on korvaavana tietona käytetty tuntikohtaista kiitotiedataa, joka antaa



Kuvio 16. Oulusta lähtevän liikenteen rullausaikojen ennusteet (N=2203), $\rho_{X,Y}$= 0,753

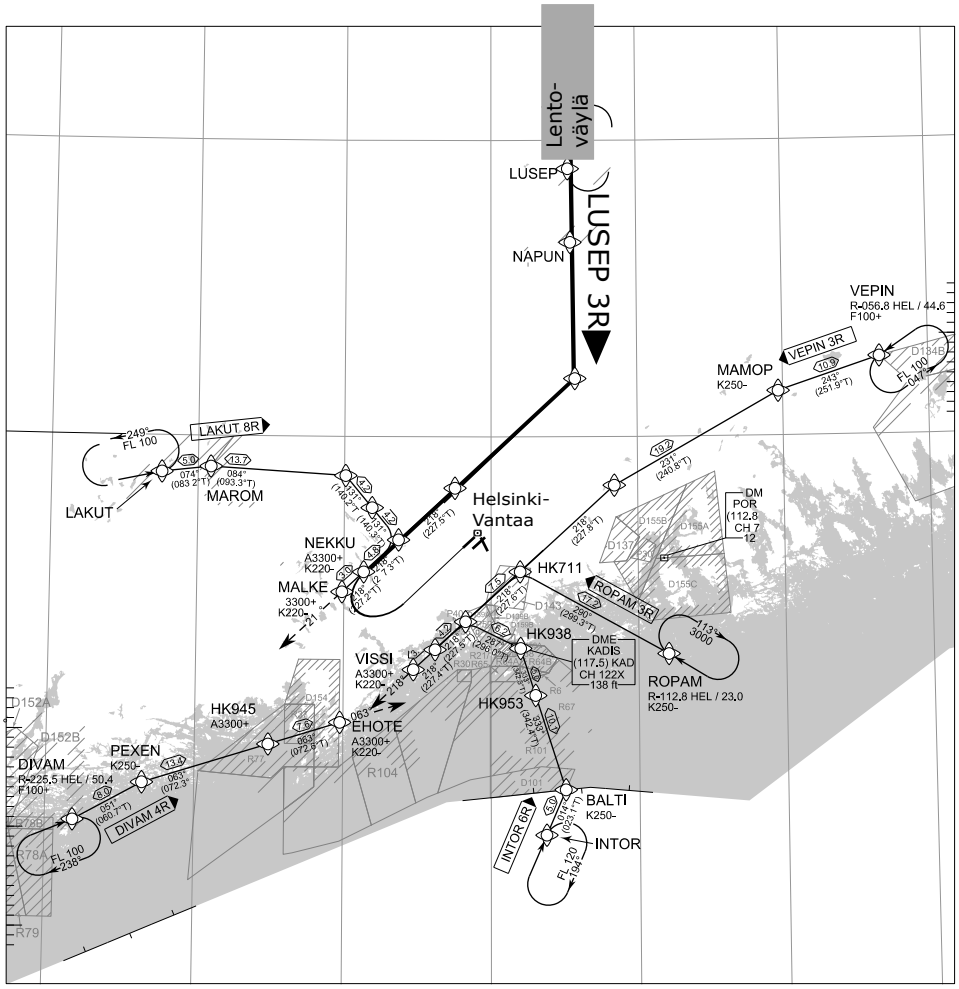
yleisen kullekin tunnille olleen kiitotien. Syötemuuttujien merkittävyydet lopulta osoittivat, että tuntiperusteinen kiitotie-ennuste oli käyttökelpoinen tämän vaiheen ennustamiseen.

4.3 Lentoaika

Lennonsuunnittelussa laskettavat lentoajan laskelmat perustuvat muutama tunti ennen lentoa tehtyihin laskelmiin, jotka perustuvat sääntöpohjaisiin menetelmiin ja ennustettuihin yläilmakehän tuulitietoihin. Lentoaika olisi laskettavissa kyseisillä suunnitteluohjelmistoilla ja tuuliennusteilla, mutta tässä tutkimuksessa pyritään lentoaika ennustamaan koneoppimisen menetelmällä kuten muutkin vaiheet.

Ennusteen toinen vaihe on reittilento, joka alkaa, kun lentokone aloittaa lentoonlähdön kiitotieltä ja nousee vakiolähtöreittiä lentoväylään. Lähtöreitillä voi olla vähäisen liikenteen aikana oikaisuja suoraan lentoväylälle. Tuloreittiäkin voidaan lyhentää oikaisuilla liikennetilanteen salliessa. Lähtöreitti, reittilento ja tuloreitti ovat laskettava erikseen toisen vaiheen kokonaisuuteen. Lentokoneen kaartamisesta reitillä aiheutuva häviö on jätetty huomioimatta

nousuun ja liukuun. Mikäli lähdön ja tulon profiili lennetään optimilla profiililla, muutokset lennon pituudessa voidaan olettaa vaikuttavan reitin vaakalentovaiheen pituuteen. Helsinki–Vantaalle pohjoisesta lentoväylää pitkin saapuva lento jatkaa vakiotuloreittiä LUSEP 3R kentän läheisyyteen, josta liitytään lähestymismenetelmään laskua varten. Tuloreitti on kuvassa 17. Tuloreiteissä voi olla korkeusrajoituksia, mutta muutoin se pyritään lentämään jatkuvan liu'un menetelmällä moottorit lentotyhjäkäynnillä.



Kuvio 17. Helsinkiin kiitotielle 04R saapuvan liikenteen vakiotuloreitit

Koska Oulussa ei tiedetä käytössä olevaa kiitotietä, voidaan kiitotie yrittää ennustaa sääntöpohjaisesti. Tulokiitotien määrittämiseen on käytetty sääntöpohjaista menetelmää, jonka käytettävyys todetaan menetelmän antaman merkityksellisyyden perusteella. Sääntöpohjai-

nen arvio perustuu ensisijaiseen kiitotiehen, jonka käyttö on edullista sekä ajallisesti että polttoainetaloudellisesti. Tuota ensisijaista kiitotietä käytetään aina sään salliessa. Ensisijaisen kiitotien käytössä on kaksi mahdollisuutta: näkölähestyminen suoraan loppuosalle tai mittarilähestymismenetelmän mukaisesti tuloreittiä pitkin. Nuo kaksi vaihtoehtoa ovat samat myös toissijaiselle kiitotielle. Näkölähestyminen on aina lyhyempi vaihtoehto kuin mittarilähestyminen. Näkölähestyminen vaatii sääolosuhteet, jotka sallivat lentokoneen laskeutumisen riittävän aikaisin pilven alapuolelle riittävän hyvään näkyvyyteen. Laskeutumisen rajana tulee käyttää lähestymissektorin minimilaskeutumiskorkeutta lisättynä käytännön lisämarginaalilla. Hyvän näkyvyyden rajaksi on käytetty näkölentosääntöjen mukaista näkyvyyttä. Pilven alarajan käyttö rajataan määräävään pilvisyyteen sääsanomassa. Mikäli joudutaan käyttämään ensisijaiselle kiitotielle epätarkkaa menetelmää, näkyvyydelle ja pilven alarajalle käytetään menetelmän mukaisia rajoituksia lisättynä pienellä varmuuslisällä. Oulussa suositaan lähestymisessä kiitotietä 30, koska se on lyhyin lentomatkaltaan. Ajallisesti kentän toiseen päähän käyttö lisää lentoaikaa, mutta rullauksessa säästyy vastaavasti rullausaikaa. Lentoajan lyheneminen kuitenkin säästää lentokuluja. Näillä perusteilla tehdään sääntö, joka määrittää kiitotien, jota olisi edullista käyttää lähestymisen aikaisissa olosuhteissa.

Lentoaika määräytyy pääasiassa konetyypin matkanopeudesta, lentoyhtiön lentomenetelmistä ja ylätuulesta. Syötemuuttujina on käytetty muutettua pintatuulta lähtö- ja tulokentällä. Muutettuna tuulena on käytetty kaksinkertaista pintatuulen voimakkuutta ja suuntaa, joka on käännetty 30 astetta ilmoitetusta tuulen suunnasta. Tämä muutettu tuuli aiheutuu maanpinnan kitkan vaikutuksesta, joka hidastaa pintatuulta ja kääntää sitä. Lisäksi on käytetty syötteenä Jyväskylän ja Oulun kohdalla olevaa ylätuulta. Näitä ylätuulia on valittu kahdelta korkeudelta. Korkeudet ovat lentopinta 180 ja 340.

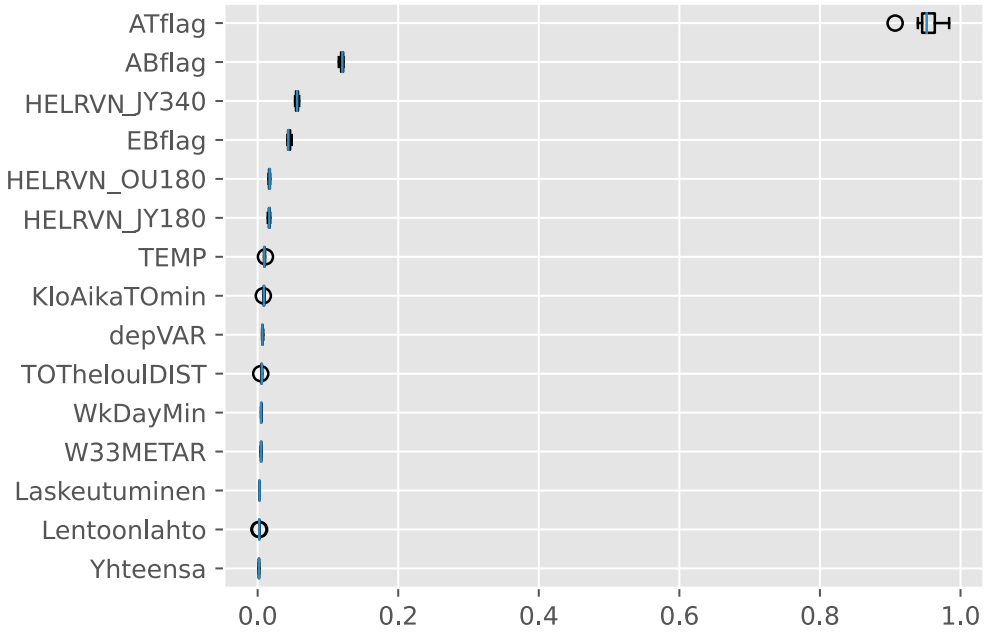
Meno-paluulennoilla paluussa tuulen vaikutus lentoaikaan on päinvastainen, mutta ei täysin sama ajallisesti. Meno-paluulennolla tulee tuulen vaikutuksesta ajallista lisää kokonaisaikaan. Lisäksi menomatkalla voitettua aikaa ei voida hyödyntää paluuseen, koska aikataulu määrää paluun lähtöajan. Ruuhka-aika voi aiheuttaa tuloaikaan viivästystä, joka saadaan aikaiseksi tuloreittiä pidentämällä tai vauhtia hidastamalla. Hidastus on polttoainetaloudellisempaa ja se on mahdollista, jos tiedetään ajoissa viivytyksen määrä. Lentoyhtiö lentää aina kokonaistaloudellisella lentonopeudella eli poikkeama lentomenetelmän nopeudesta suun-

taan tai toiseen aiheuttaa lisäkuluja. Siksi olisi tärkeää pystyä ennustamaan lentoaika, mikä auttaisi saavuttamaan optimin lähestymisprofiilin.

METAR:sta tuodaan lennon tietoihin lähtö- ja tulokentän säätiedot arvioidun lähtö- ja laskeutumisajan mukaisesti. Lennon tiedoille on tuotava kahdet säätiedot, joita tarvitaan lähtö- ja tulokentän kiitotien todennäköiseen valintaan.

Lentoajan selvästi merkittävin tekijä on lentokoneen tyyppi. Seuraavana vaikuttavana syötemuuttujana on lentoväylän tuuli Jyväskylän kohdalla lentopinnalla 340, joka vastaa suihkukoneiden käyttämää lentokorkeutta. Toinen vastaava tuuli potkurikoneille on lentopinnan 180 tuuli Jyväskylän sekä Oulun kohdalla. Helsingin kohdalla olevia ylätuulia ei ole datassa. Muuten syötemuuttujat ovat vähemmän merkittäviä. Huomattavaa on, että lennon pituus tulee vasta lämpötilan, päivänajan ja lähtöviivästyksen jälkeen. Vain kahden syötemuuttujan tiedot tulevat METAR- tai TAF-sanomasta. Syötteitä on 15 kappaletta. Toisessa iteraatiossa voidaan vielä poistaa joitain syötetietoja käytöstä.

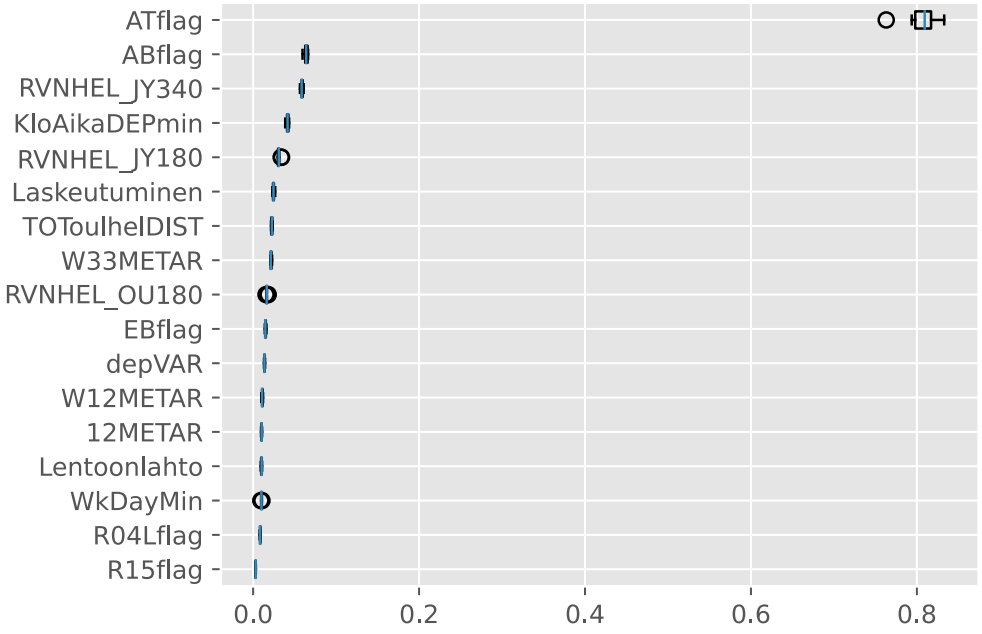
Permutaation merkityksellisyydet voidaan lukea kuvissa 18 ja 19.



Kuvio 18. Helsinki–Oulu lentoaikojen syötemuuttujien merkittävyydet

Oulusta Helsinkiin lennettäessä merkittävät syötemuuttujat eivät ole täysin samoja. Päiván-

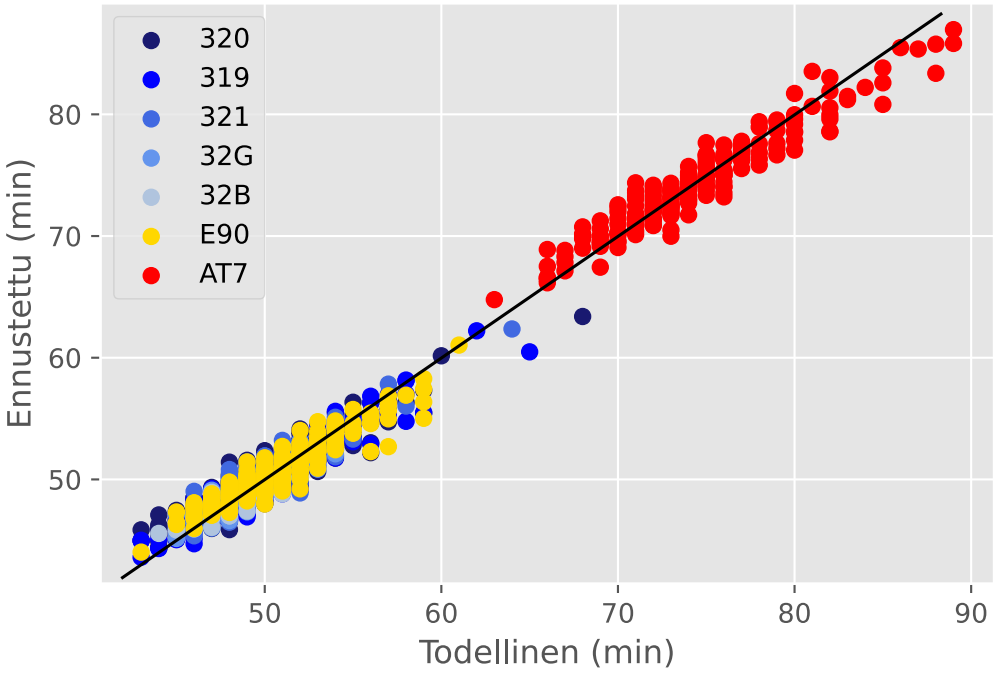
ajan hetki korostuu, johtuen mahdollisesti tuloajan ruuhkista Helsingissä. Samaan viittaa Helsingin laskeutuvan liikenteen määrän merkitys. Oulun ja Jyväskylän kohdalla olevalla alemmalla ylätuulella on hieman suurempi merkitys kuin pohjoiseen mentäessä. Käytetyt syötemuuttujat ovat lähes kaikki rikastettuja ja niiden lukumäärä oli 2–3 kertainen alunperin testattuihin syötteisiin verrattuna.



Kuvio 19. Oulu–Helsinki lentoaikojen syötemuuttujien merkittävyydet

Molemmilla reiteillä lentoaika muodostaa ajallisesti ison osan kokonaisajasta ja molempiin suuntiin suurin tekijä on se, onko konetyyppi ATR. Pelkän konetyypin ja ylätuulen perusteella olisi todennäköisesti mahdollista tehdä nopea ennuste, josta saadaan ennakkovaroitus mahdollisesta lennon viivästymisestä. Seuraavassa kuvaajassa 20 on kuvattu Helsinki–Oulu välin lentoajan ja ennustetun lentoajan suhteet. Ennusteen keskimääräinen virhe Ouluun on 50 sekuntia. Tulos on korrelaation 0,991 mukaan erittäin tarkka ja ennusteeseen vaikuttaa pääasiassa konetyyppi ja kahden korkeuden ylätuulikomponentit. Tässä ennusteessa ovat mukana oletukset lähtö- ja tuloreitistä sekä lennon aikaiset muun liikenteen aiheuttamat poikkeukset. Mahdollisesti lennonvalmistelussa tehdyt viranomaismääräykset täyttävät laskelmat eivät ole yhtä tarkkoja.

Kuviossa 21 on kuvattu Oulu–Helsinki välin todellisen ja ennustetun lentoajan suhteet. En-



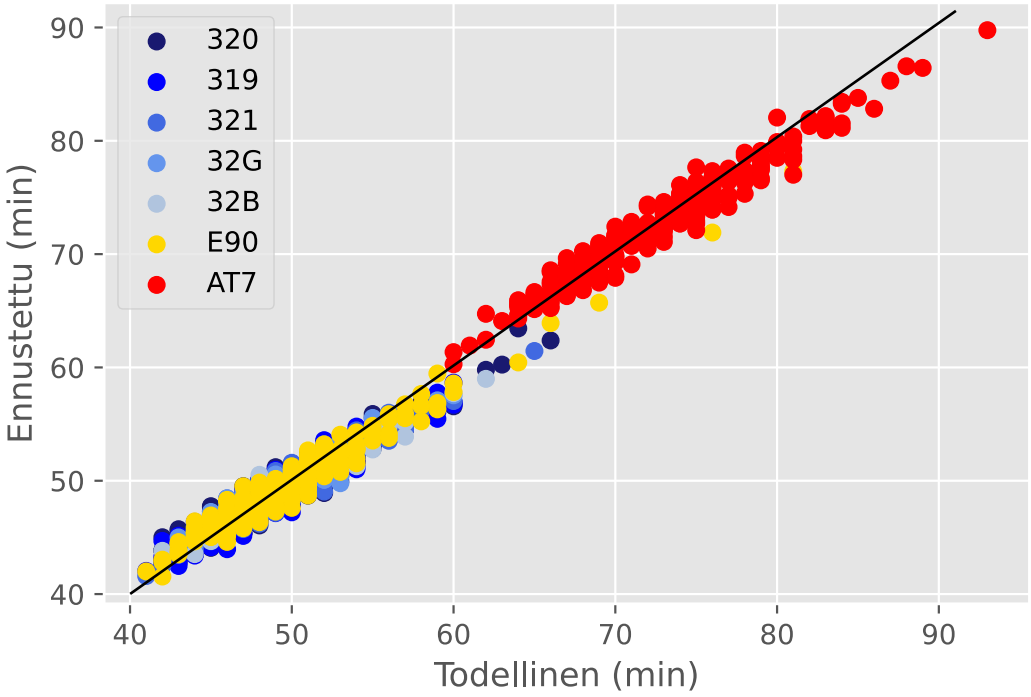
Kuvio 20. Helsinki–Oulu lentoaikojen ennusteet (N=1832), $\rho_{X,Y}$= 0,991

nusteen keskimääräinen virhe Helsinkiin on 53 sekuntia ja korrelaatio 0,993. Lentoaikojen ennusteet ovat molemmille kentille erittäin tarkkoja. Molemmissa on konetyypillä iso vaikutus tulokseen. Lentoajoissa on suuri ajallinen vaihtelu 40–90 minuutin välille ja pelkästään konetyyppi kohtainen vaihtelukin on noin 30 minuuttia. Helsingin suuntaan vaihtelu on isompaa eli poikkeuksia on enemmän, mutta ne ovat ennustettavissa.

Lentovaihe päättyy määräkentän kiitotielle, josta alkaa seuraava vaihe. Lentovaiheen jälkeen voidaan ennusteesta jo laskea, kauanko laskeutuneella lentokoneella on aikaa saapua aikataulun mukaisesti portille. Näillä aikatauluun suhteellisilla ajoilla on inhimillinen vaikutus esimerkiksi rullaukseen käytettävään aikaan. Inhimillisen tekijän vaikutusta voidaan arvioida seuraavan alaluvun ennusteesta ja syötemuuttujien permutaation vaikutuksista.

4.4 Rullausaika kiitotieltä tuloportille

Tämän alaluvun vaiheessa vaikuttavat tekijät rajautuvat laskukiitotien ja tuloportin välille. Laskukiitotie ja porttitiedot puuttuvat Oulun osalta, mikä on huomattava haaste. Laskukiitotie määrää, kuinka pitkä matka rullataan portille ja onko rullauksen aikana ylitettäviä kiito-

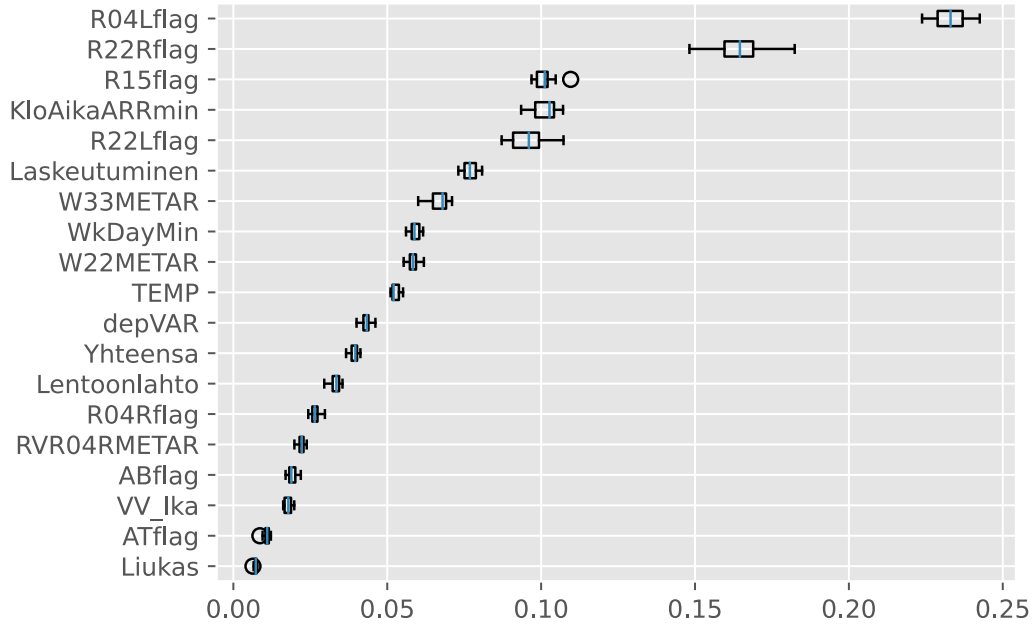


Kuvio 21. Oulu–Helsinki lentoaikojen ennusteet (N=2227), $\rho_{X,Y}$= 0,992

teitä ja niiden aiheuttamia rajoitteita. Eli käytännössä samat tekijät vaikuttavat kuin lähtiessä, mutta jäänpoiston osalta tilanne on erilainen.

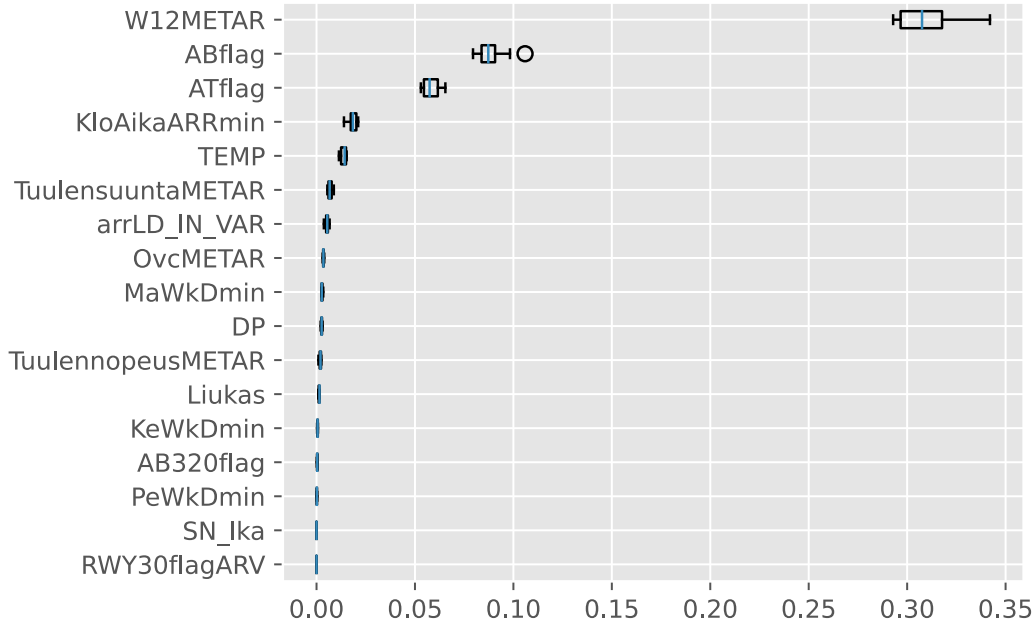
Helsingin laskun jälkeen rullaavissa lennoissa ovat mukana opetuksessa myös Rovaniemeltä tulevat lennot. Näin saadaan paljon lisää tapahtumia ennusteen opetukseen. Vain kolme syötemuuttujaa oli säähän liittyviä. Helsingissä syötemuuttujien merkityksellisyyksissä korostuvat käytössä oleva kiitotie ja liikennemäärät. Seuraavaksi eniten vaikuttavat kiitotien tuulitiedot. Säällä ei ole isoa merkitystä, mutta huonon sään menetelmien mukaiset sääsyötteet ovat merkittäviä. Kuviossa 22 on kuvattu Helsingin laskun jälkeisen rullausajan syötemuuttujien merkityksellisyyttä.

Oulun osalta syötemuuttujissa korostuvat Helsinkiä enemmän olosuhteet ja konetyypit. Kiitotietiedot puuttuivat Oulun aineistosta. Oulun syötemuuttujista vaikuttavin on tuuli (0,31). Seuraavaksi eniten vaikuttaa konetyyppi. Nämä ovat merkittävimmät tekijät ja muiden osuus on kohtuullisen pieni. Mikäli menetelmää yksinkertaistettaisiin, voitaisiin käyttää pelkkiä konetyyppejä ja kiitotien tuulitietoa rullausajan ennustamiseen. Helsinkiin verrattuna merkittävien tekijöiden osuus on pieni. Kahdeksan syötettä oli säähän liittyviä. Oulussa yli seit-



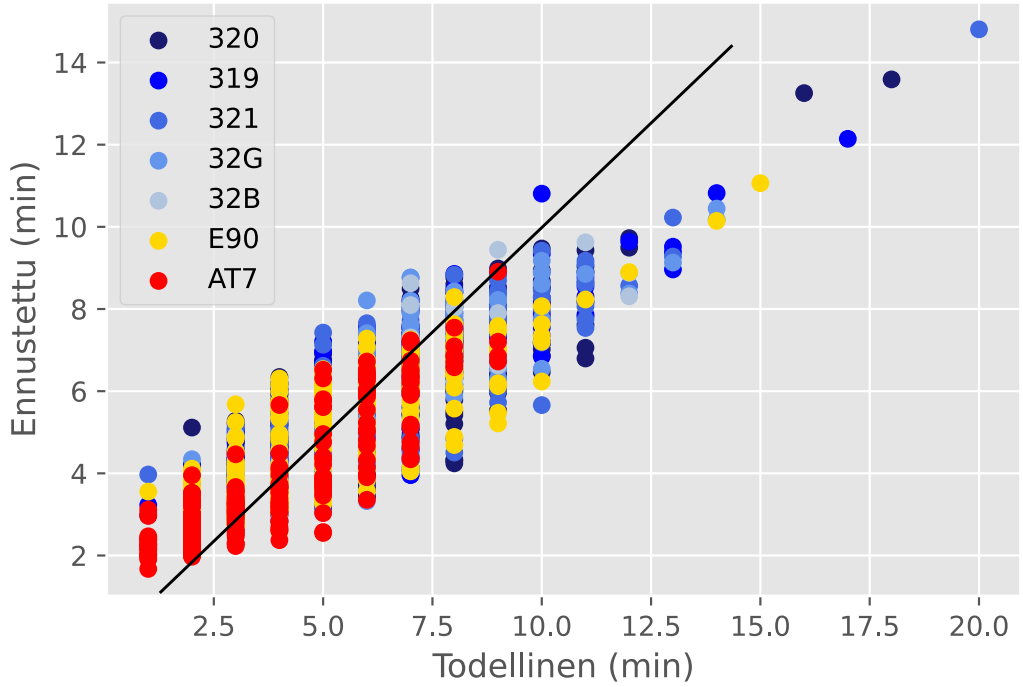
Kuvio 22. Helsinkiin tulevan liikenteen rullausaikojen syötemuuttujien merkityksellisyys

semän minuutin rullausajat ovat ääriarvoja ja ne ovat poistettu, jotta ennuste ei muuta tavanomaisen liikenteen arvoja. Neljä lentoa 2044 lennosta ylitti tuon rajan. Kuviossa 23 on kuvattu Oulun laskun jälkeisen rullausajan merkitykselliset syötemuuttujat.



Kuvio 23. Ouluun tulevan liikenteen rullausaikojen syötemuuttujien merkityksellisyys

Laskun jälkeisen rullausajan ennuste Helsingin lentoaseman tuloportille on kuviossa 24. Ennusteen keskimääräinen virhe Helsingissä on 49 sekuntia. Huolimatta lentokentän monimutkaisuudesta, ennusteen korrelaatio on hyvin vahva (0,88). Joitain syötemuuttujia voitaisiin vielä lisätä toiseen iteraatioon ja parantaa tarkkuutta kuten lähtörullauksissa.

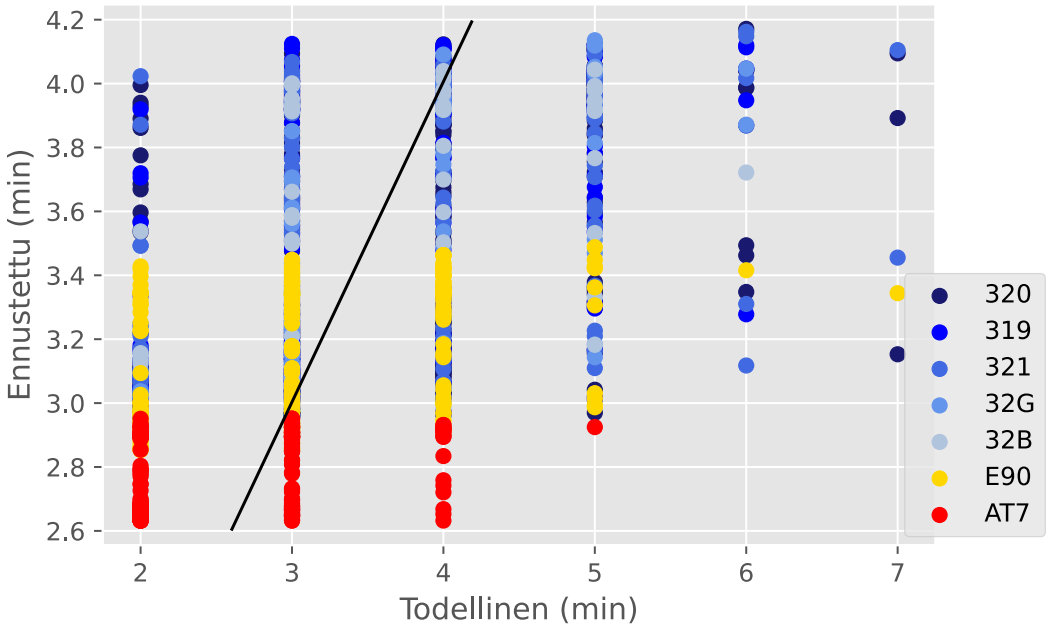


Kuvio 24. Helsinkiin tulevan liikenteen rullausaikojen ennusteet (N=3491), $\rho_{X,Y}$= 0,880

Ennusteen keskimääräinen virhe Oulussa on 52 sekuntia. Ajallisesti rullausaikojen ennusteessa Oulun keskimääräinen virhe on samaa luokkaa kuin Helsingin, vaikka Helsingin rullausajat ovat kolme kertaa pidempiä. Oulun ennusteiden suhteellinen virhe on kuitenkin huomattava, mikä näkyy korrelaatiossa (0,585). Rullaukset päättyvät portille tai seisontapaikoille, joiden sijaintia ei tiedetä. Tästä johtuen rullausmatkan arviointi on vaikea saada tarkaksi.

Kuviossa 25 on kuvattu laskun jälkeisen rullausajan ennuste Oulun lentoaseman tuloportille. Ennusteet eivät täysin vastaa todellisia aikoja pidemmillä rullausajoilla. Pidempiä rullausaikoja voitaisiin tutkia ja pyrkiä löytämään niille yhteinen tekijä, joka voitaisiin lisätä syötemuuttujiin. Tämä ennuste vaatisi lisää analysointia hyvän syötteen löytämiseksi.

Tähän asti lennon vaiheita yhdistävänä avaintekijänä on ollut lennon numero. Samana päivänä on vain yhdellä lennolla yksilöllinen lennon numero, joka on sama kaikissa lennon



Kuvio 25. Ouluun tulevan liikenteen rullausaikojen ennusteet (N=2040), $\rho_{X,Y} = 0{,}585$

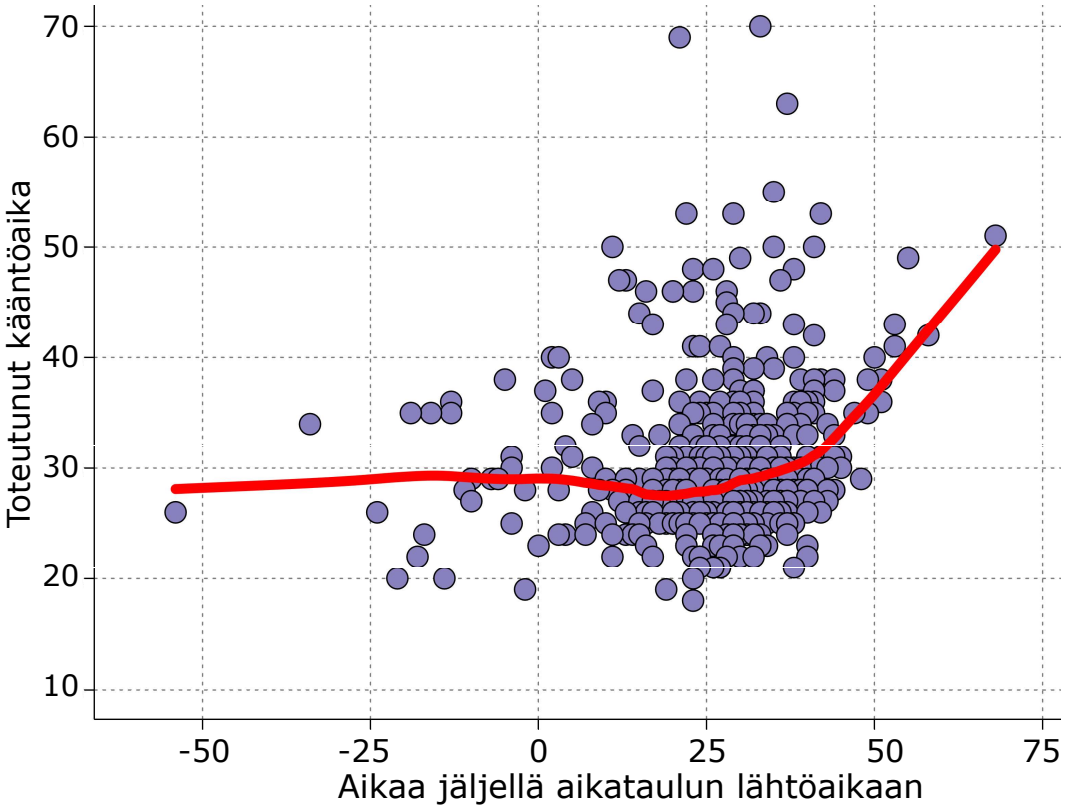
vaiheissa. Tätä lennon numeroa käytetään kyseisen lennon vaiheiden yhteenlaskemiseen. Seuraavassa vaiheessa eli kääntöajassa lennon numero vaihtuu. Tämän haasteen ratkaisua selitetään tarkemmin seuraavassa alaluvussa.

4.5 Kääntöaika

Portille saapumisen jälkeen alkaa kääntöaika. Tämän vaiheen avaintekijänä on lentokoneen rekisterinumero, joka säilyy samana, vaikka lennon numero vaihtuu. Saman lentokoneen tulo- ja lähtöaikojen vertailuun tehdään aputaulukko, jossa on kunkin lentokoneen saapumisajat Ouluun eri päivinä. Koneen ollessa lähdössä Oulusta haetaan aputaulukosta lähin tuloaika menneisyydestä saman päivän aikana. Tämän haetun tuloajan ja lennon lähtöajan väli on kääntöaika, joka toimii koneoppimisen syötemuuttujana kääntöaikaa opetettaessa. Uusi syötemuuttuja on aika portille tulosta aikataulun mukaiseen lähtöaikaan.

Konetyyppikohtaisesti vallitsevissa olosuhteissa on olemassa minimikääntöaika, jossa kone voidaan kääntää. Kääntöaika noudattelee sääntöä, joka on riippuvainen lähtöön jäljellä olevasta ajasta ja minimikääntöajasta. Säännön muodostamisessa tukeudutaan koneoppimiseen.

Kuvassa 26 on pystyakselilla toteutunut kääntöaika. Vaaka-akselilla on jäljellä oleva aika aikataulun mukaiseen lähtöaikaan. Jos kääntöaika on negatiivinen, aikataulun mukainen aika on jo ohitettu. Siinä tapauksessa lentokoneen kääntöajaksi tulee minimikääntöaika. Kuvassa Embraer-kone käännetään keskimäärin 29 minuutissa.



Kuvio 26. Oulun toteutunut kääntöaika suhteessa aikataulun mukaiseen käytössä olevaan aikaan

Jos ollaan lähellä tuota minimiä kääntöaikaa, pyritään erityisen nopeaan kääntöön, jottei tule merkintää myöhästymisestä. Jos aikaa jää enemmän kuin normaalin kääntöajan verran, ei voida lähteä etuajassa. Tällöin odotellaan viimeistä matkustajaa tai lähtöaikaa ja kääntöaika muodostuu normaalia kääntöaikaa pidemmäksi. Kuvion pisteistä nähdään, että kääntöaikaan vaikuttaa muitakin tekijöitä, jotka tekevät vaihtelua tähän keskimääräiseen sääntöön. Koneoppimisella ennustetaan tällaisesta monen syötemuuttujan vaikutuksesta vastausta toteutuvaksi kääntöajaksi.

Tutkimuksessa ei ennusteta seuraavan lähtöajan myöhästymistä. Sen sijaan ennustetaan, koneen kääntöaikaa portilta. Mikäli koneoppiminen kykenee edellä mainitun säännön luomi-

seen, saadaan todennäköinen kääntöaika näissä olosuhteissa ja siitä saadaan lähtöajan ennuste. Tuosta ennusteesta nähdään haluttaessa, tuleeko lento olemaan myöhässä.

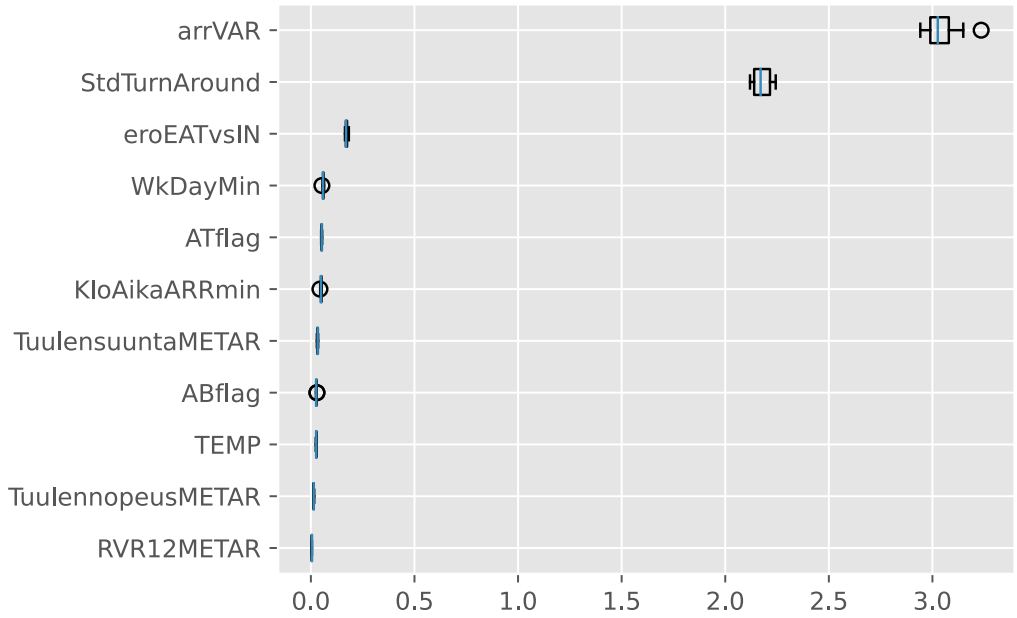
Aiempien tutkimusten mukaan matkustajamäärällä ja rahdin määrällä on iso vaikutus minimikääntöaikaan. Jotta tämä tutkimus olisi julkinen, matkustajamäärät ja rahdin määrät ovat jätetty pois aineistosta. Seuraavassa iteraatiossa on näille tärkeille syötemuuttujille pyrittävä löytämään korreloiva tekijä ja tetata sen vaikutusta ennusteeseen.

Kääntöaikaan vaikuttavat myös ruuhkat kentällä sekä lentokoneen seisontapaikan sijainti. Myöskin bussikuljetukset, maahenkilöstön saavutettavuus, yhteydet asemahenkilöstöön, huolintayhtiön sopimukset, jäänpoisto ovat tekijöitä, joita on huomioitava syötemuuttujissa. Kuitenkaan lento ei yleensä voi lähteä etuajassa ilmoitetusta lähtöajasta.

Kahdesta lennosta muodostuvan meno-paluulennon ennustaminen on mahdollista ketjuttamalla lentokenttien tunnettujen lentopareja. Ketjuttamiseen vaikuttaminen edellyttää lentokoneiden sekä lentohenkilöstön suunnitellun työsarjan tietämistä (pairings ja roster). Näitä pareja voidaan muokata lentoyhtiössä, etenkin henkilöstön ja konetyyppien osalta, jotta poikkeamiin voitaisiin vaikuttaa. Seuraavan päivän sääennusteilla voisi silloin ennustaa, kuinka tullaan onnistumaan suunnitellussa aikataulussa. Mikäli ennustetaan poikkeamaa, voitaisiin muuttaa suunnitelmaa myöhästymisen välttämiseksi.

Mikäli kääntöaika on pidempi kuin aikatauluun jäljellä oleva aika, lento tulee myöhästymään. Myös kääntöaikaa pidemmät käytössä olevat kääntöajat voivat myöhästyä, kuten yöpymisen yhteydessä olevat useamman tunnin käännöt. Nämä tapaukset tulee myös huomioida syötemuuttujien valinnassa. Vaikka lumisade ei vaikuta normaaliin kääntöaikaan, se voi olla ongelma aamun miehistökuljetukselle ja lentokentälle tulevalle matkustajille. Ensimmäisessä iteraatiossa ei aamun myöhästymisten määrää ole ennustettu yön kääntöaikaan.

Kuviossa 27 on kuvattu Oulun kääntöaikaan vaikuttavien syötemuuttujien merkityksiä. Selkeästi merkittävin syötemuuttuja on portille tulon poikkeama normaalista tuloajasta, jonka vaikutus on 3,0. Toinen merkittävä tekijä on aikataulun mukaiseen kääntöaikaan jäljellä oleva aika. Kääntöaikaa voitaisiin ennustaa hyvin vähäisillä syötemuuttujilla. Kolmas syötemuuttuja oli laskeutumisen todellisen ajan ja aikataulun mukaisen portille tulon välinen aika. Portille tulon poikkeama voidaan laskea edellisten ennustevaiheiden jälkeen.

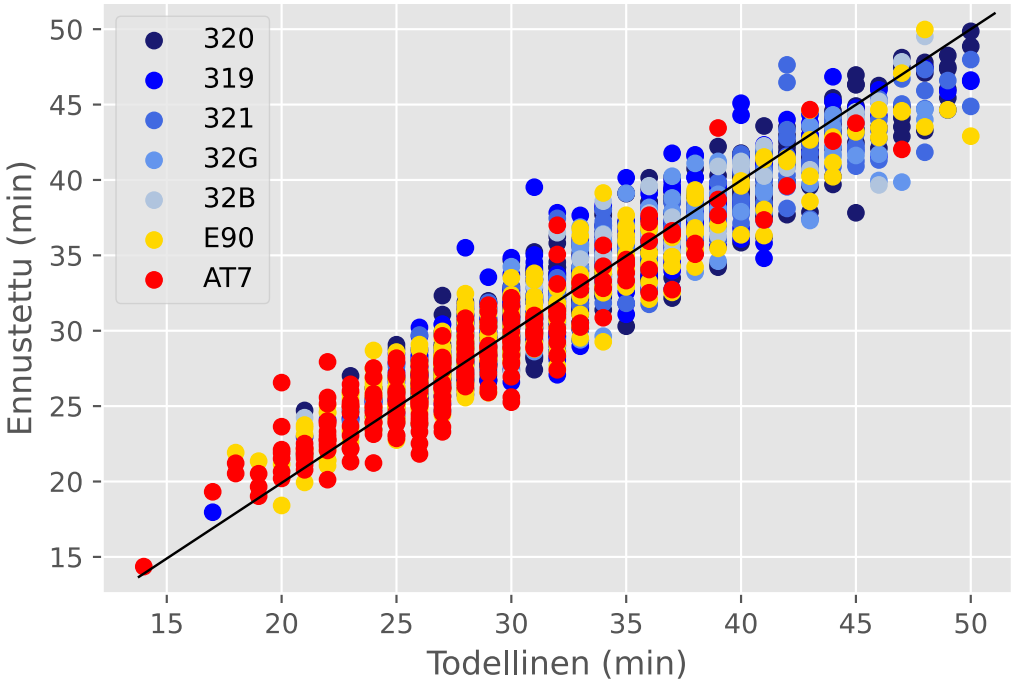


Kuvio 27. Oulun kääntöajan syötemuuttujien merkitykset

Helsingin kääntöaikoja ei voida ennustaa, koska lentoyhtiön mukaan kotimaan lennon jälkeen tulee yleensä seuraavana ulkomaan lento. Aineistossa on pelkkiä kotimaan lentoja. Myöskään Helsingistä lähtevälle koneelle ei voida ennustaa kääntöajasta johtuvaa myöhästymistä, koska lentotiedoissa ei ollut lentokoneen edelliseltä lennolta saapumisaikaa. Oulun kääntöajan ennuste kuitenkin osoittaa, että kääntöajan ennuste olisi ennustettavissa ja tarpeellinen Helsingissäkin, kunhan saadaan siihen tarvittavat tiedot. Kuviossa 28 on kuvattu kääntöaikaa Oulun lentoasemalla.

Ennusteen keskimääräinen virhe on yksi minuutti 32 sekuntia. Ennusteen korrelaatio on erittäin vahva (0,953). Kuvaajan mukaan hajonta on pieni koko aikavälillä ja kaikilla konetyypeillä.

Nyt ovat kaikki lennon vaiheet ennustettu ja niistä voidaan koota meno-paluulennon kokonaisennuste. Seuraavassa alaluvussa lasketaan yhteen seitsemän osaennustetta ja todetaan summautuvatko virheet liian suuriksi, vai ovatko virheet eri suuntaisia ja nollaavat toisensa.

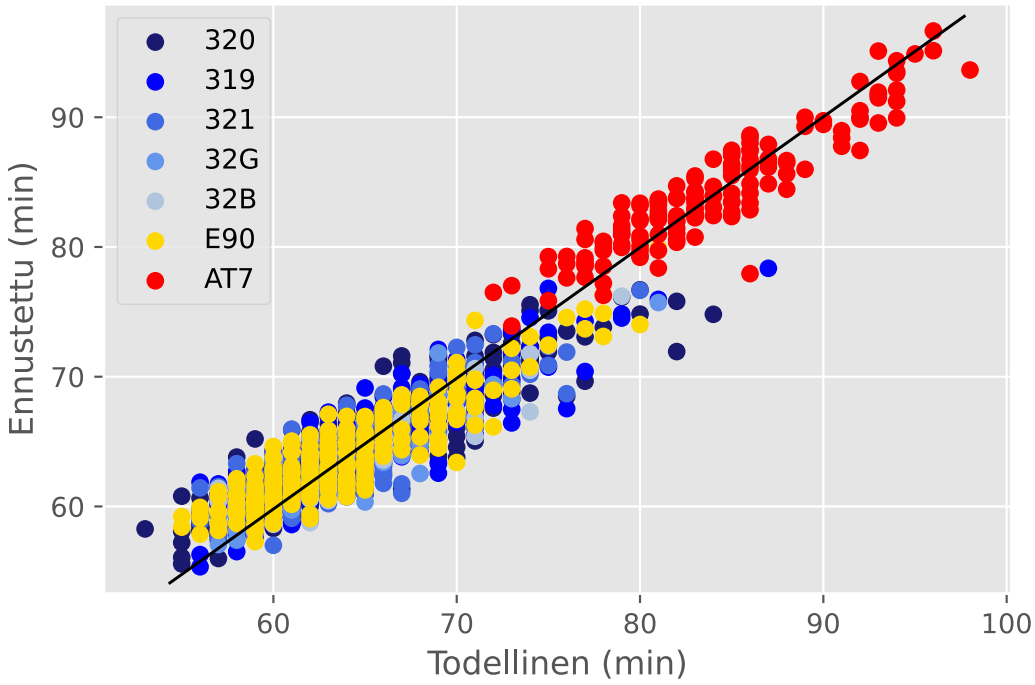


Kuvio 28. Oulun kääntöajan ennusteet (N=1705), $\rho_{X,Y}$= 0,953

4.6 Mallien vaiheiden yhdistäminen

Kokonaisennusteella tutkitaan, kuinka tarkka on opetusdatalla tehty ennuste nykyisillä syötemuuttujilla. Aluksi vertaillaan yhden lennon koonnoksen tarkkuutta sekä meno- että paluulennolta. Yhden lennon koonnos saadaan summaamalla lennon vaiheet yhteen. Kun todelliset ja ennustetut ajat ovat summattu, niitä vertaillaan kuvaajassa ja lasketaan MAE sekä korrelaatio koosteelle. Kuvaajan vaaka-akselilla on lennon vaiheiden toteutuneiden aikojen summa ja pystyakselilla samalle lennolle tehtyjen ennusteiden summa. Lisäksi esitetään tarkempia kuvaajia ennusteen ja todellisen ajan erosta. Ennusteen ja todellisen ajan kuvaajasta nähdään, jos ennusteissa on säännönmukaisia poikkeamia. Summauksessa ei tehdä kokonaisvirheen optimointia, vaan virheet ovat minimoitu vaiheittain.

Aika Helsingin lähtöportilta Oulun tuloportille sisältää rullausajat lähtö- ja määräkentillä sekä niiden välisen lentoajan. Kokonaisvirhe ei summaudu vaiheista, koska vaiheiden virheet kumoavat satunnaisesti toisiaan aikoja yhdistettäessä. Kokonaisennusteen keskimääräinen virhe Helsingistä Ouluun on yksi minuutti 43 sekuntia ja korrelaatio 0,991. Kuviossa 29 on kuvattu Helsinki–Oulu välin kokonaisaika ja -ennuste.

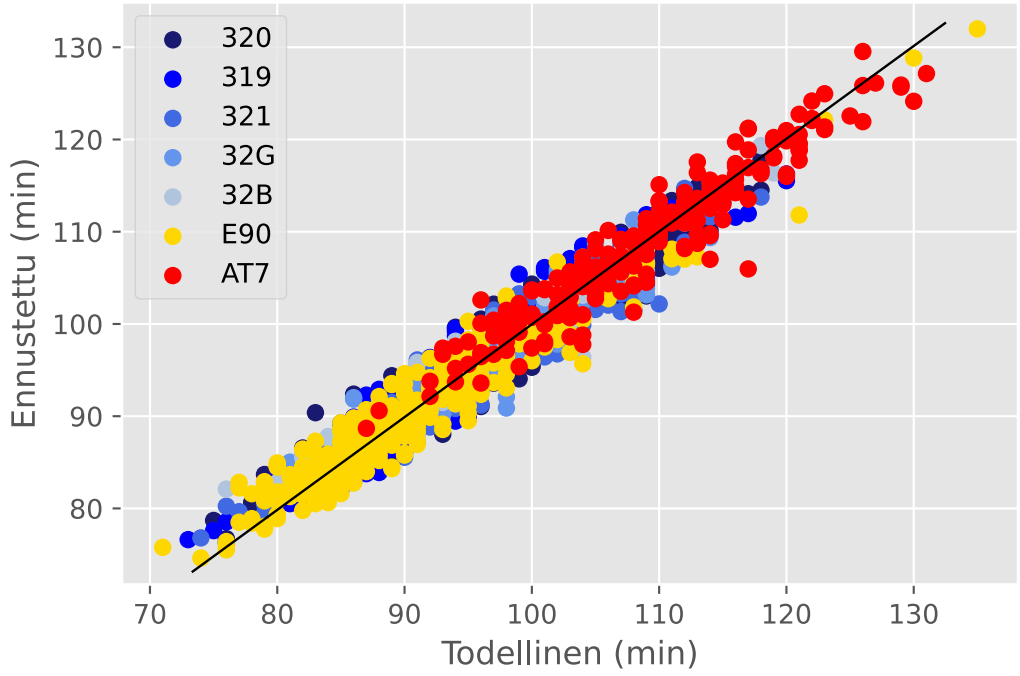


Kuvio 29. Helsingin lähtöportilta Oulun tuloportille kuluva aika ja ennuste, $\rho_{X,Y}$ = 0,991

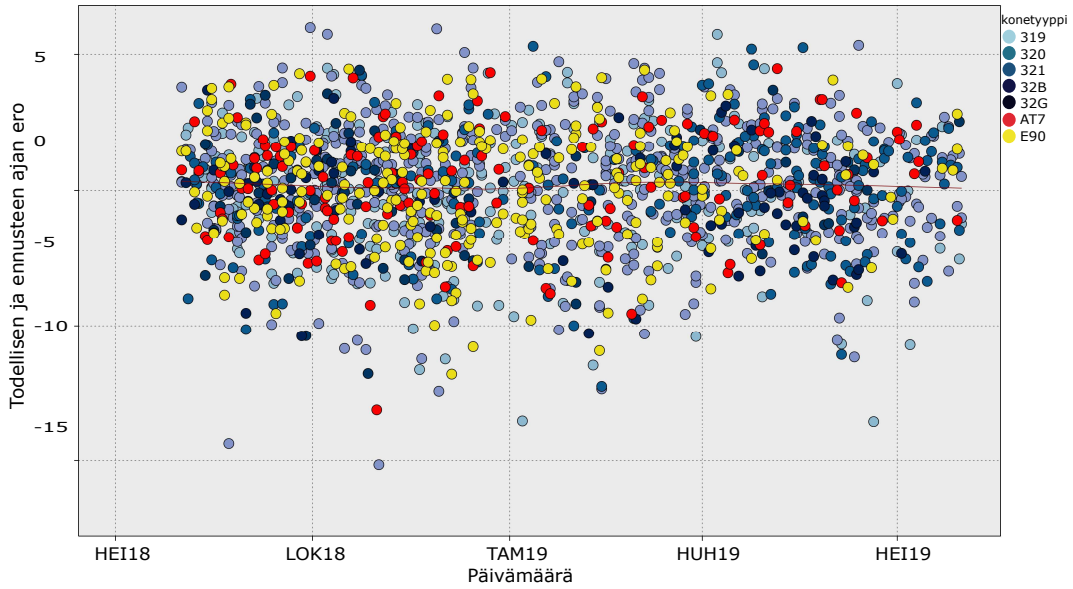
Kuviossa 30 on vastaavasti kuvattu aika ja ennuste takaisin Oulusta Helsinkiin. Tämä aika sisältää lisäksi kääntöajan Oulussa. Keskimääräinen virhe Oulusta Helsinkiin on lähes sama eli yksi minuutti 45 sekuntia ja korrelaatio 0,992. Virhe voisi olla isompi, koska aika sisältää myös kääntöajan Oulussa. Lisäksi konetyypeittäin aikojen hajonta näyttää isommalta kuin Helsinki–Oulu välillä.

Virheen suuruus eri kalenteriaikoina eri konetyypeillä selviää tarkemmin kuviosta 31 ja 31. Vaaka-akselilla on kalenteriaika ja pystyakselilla todellisen ajan ja ennusteen ero. Poikkeamat jakaantuvat molemmissa tasaisesti eri vuodenaikoina sekä eri konetyypeittäin. Ennusteissa ei näytä olevan systemaattisia virheitä.

Kuviossa 33 on kuvattu kokonaisaika ja ennuste Helsingin lähtöportilta takaisin Helsingin tuloportille. Kyseinen aika sisältää meno-paluulennon kaikkien ennusteiden summan. Keskimääräinen virhe opetusdatalla on kaksi minuuttia 32 sekuntia. Tulos on tarkkuudeltaan huomattavasti parempi kuin tavoitteena ollut viiden minuutin keskimääräinen virhe ja korrelaatio on erittäin vahva 0,974.

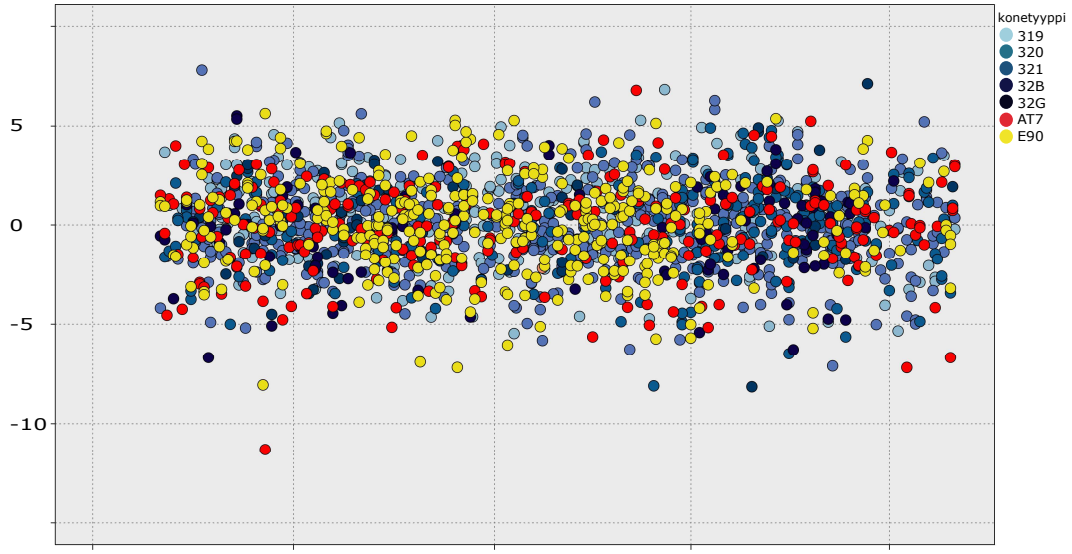


Kuvio 30. Oulun kääntöajan ja lähtöportilta Helsingin tuloportille kuluvan ajan summa verrattuna ennusteeseen, $\rho_{X,Y} = 0{,}992$

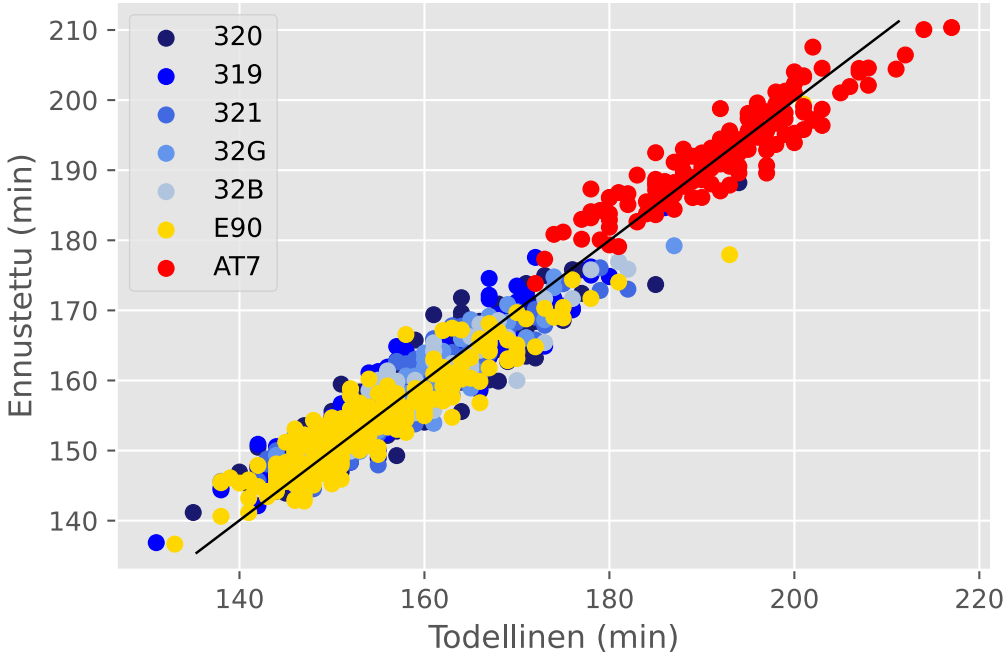


Kuvio 31. Helsingin lähtöportilta Oulun tuloportille kuluvan ajan ja ennusteen ero

Testidatalla saadaan yleensä suurempi keskimääräisen virheen kuin opetusdatalla. Mikäli 10 kansion ristiinvalidointi opetuksessa on toiminut hyvin, ylioppimista ei ole ja seuraavan



Kuvio 32. Oulun kääntöaika ja Oulun lähtöportilta Helsingin tuloportille kuluvan ajan ja ennusteen ero



Kuvio 33. Helsinki–Oulu–Helsinki lentojen aika ja ennuste (N = 1246 paria), $\rho_{X,Y}$ = 0,974

luvun vahvistamisesta voidaan olettaa tutkimuksen tavoitteen mukaisia tuloksia.

5 Ennusteiden vahvistaminen

Tässä luvussa vertaillaan kunkin vaiheen ennustetta testidatalla, minkä jälkeen ennusteet kootaan yhteen lentopareittain meno-paluulennon ennusteeksi. Kokonaisennuste testidatalla osoittaa, päästäänkö asetettuun viiden minuutin tarkkuuteen. Testidatan käytöllä varmistetaan, että menetelmä toimii myös täysin uudella tiedolla. Mikäli testidatan ennuste on merkittävästi huonompi kuin opetusdatalla, menetelmä on voinut ylioppia opetusdatan tapahtumat. Tämä ylioppiminen huonontaisi yleistyvyyttä eli tulevat uudet ennusteet eivät vastaisi uusia tapahtumia.

Testidata erotetaan aineistosta yleensä satunnaismenetelmällä. Satunaismenetelmää ei voida käyttää lennon eri vaiheita yhdistettäessä. Käytetyssä aineistossa tulee olla alusta loppuun lentoparin kaikki vaiheet eikä kokonaisuudesta voi puuttua satunnaisesti jonkin vaiheen tietoja. Testidata on siis eroteltava ennen ensimmäistä lennon vaihetta ja säilytettävä tuo jaottelu kaikissa vaiheissa. Lisäksi pitää säilyttää saman lentokoneen eri lentojen ketju. Ketjutus tapahtuu seuraamalla samaa koneyksilöä koko päivä. Tämä mahdollistaa myös kääntöaikojen ennustamisen. On siis säilytettävä koko päivältä kaikki kyseisen lentokoneen lennon vaiheet.

Tutkimuksen ennusteen vahvistaminen tehdään ennustamalla vaiheiden ajat käyttäen kyseisen hetken oikeita säätietoja. Oletuksena on siis, että sää-, kiitotie-, liikenne- ja ylätuulienusteet ovat parhaita mahdollisia. Ulkoisten ennustesyötteiden tiedoilla vahvistaminen liittyisi enemmänkin ennustesyötteiden arvioimiseen eikä sitä arvioida tässä tutkimuksessa.

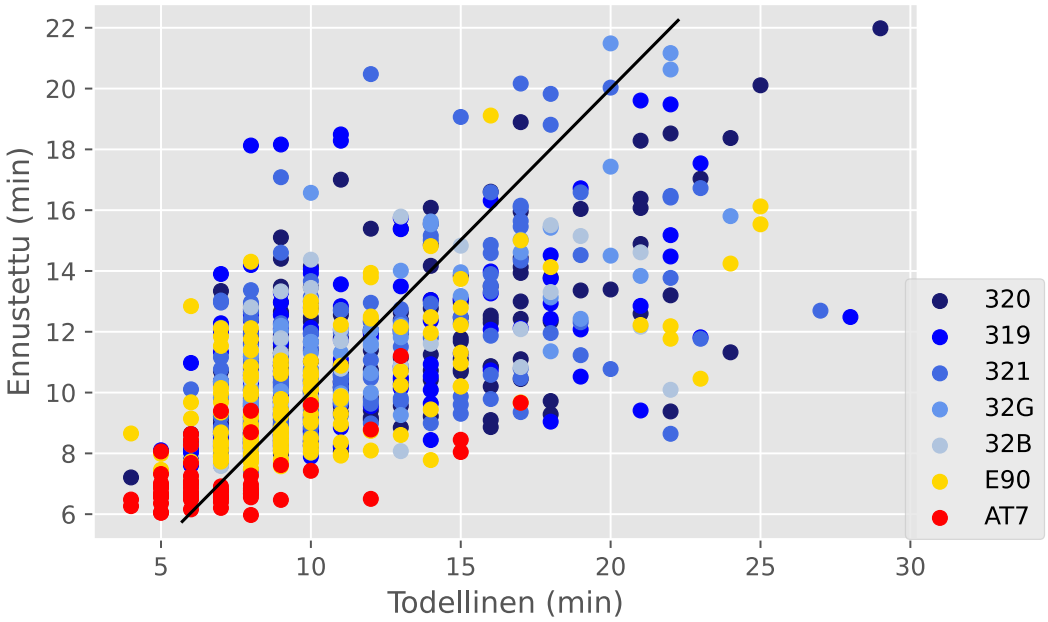
Tutkimuksessa on päädytty erittelemään testidataksi neljännes koko datasta niin, että joka neljännen päivän kaikki lennot ovat testidataa. Näin on mahdollista säilyttää lentosarjojen kokonaisuus katkeamattomana. Sarjalla saadaan kaikki vuodenajat sekä viikonpäivät tasaisesti mukaan niin, että seuraavalla viikolla testidata vaihtuu seuraavaksi viikonpäiväksi. Eli testisarjat olisivat esimerkiksi: maanantai, perjantai, tiistai, lauantai, keskiviikko, sunnuntai jne.

Summaamalla eri vaiheiden ennusteet saadaan arvio kokonaisuudesta. Ennusteiden summaa verrataan todellisten vaiheiden aikojen summaan. Vaiheiden ennusteet poikkeavat niin vähän todellisista ajoista, että todelliselle ajalle valittua säätietoa voidaan käyttää ennustamiseen.

Todellisessa tilanteessa vaiheessa käytetty aika ja sen sääennuste tulisivat edellisestä ennusteesta saadusta ajasta. Luvun lopussa on taulukko opetus- ja testidatan ennusteiden poikkeamista todellisesta ajasta vaiheittain ja eri kokonaisuuksina. Taulukossa on lisäksi ennusteen ja todellisen ajan välinen Pearsonin korrelaatio.

Mikäli sääennusteet vastaavat täysin toteutuvaa säätä, aikatauluennuste tulee olemaan yhtä tarkka kuin ennustettaessa METAR-tietojen perusteella. Myöhemmin testattavaksi jää, miten tarkka on vaiheiden kokonaisennuste TAF-sanoman tiedoilla ennustettaessa.

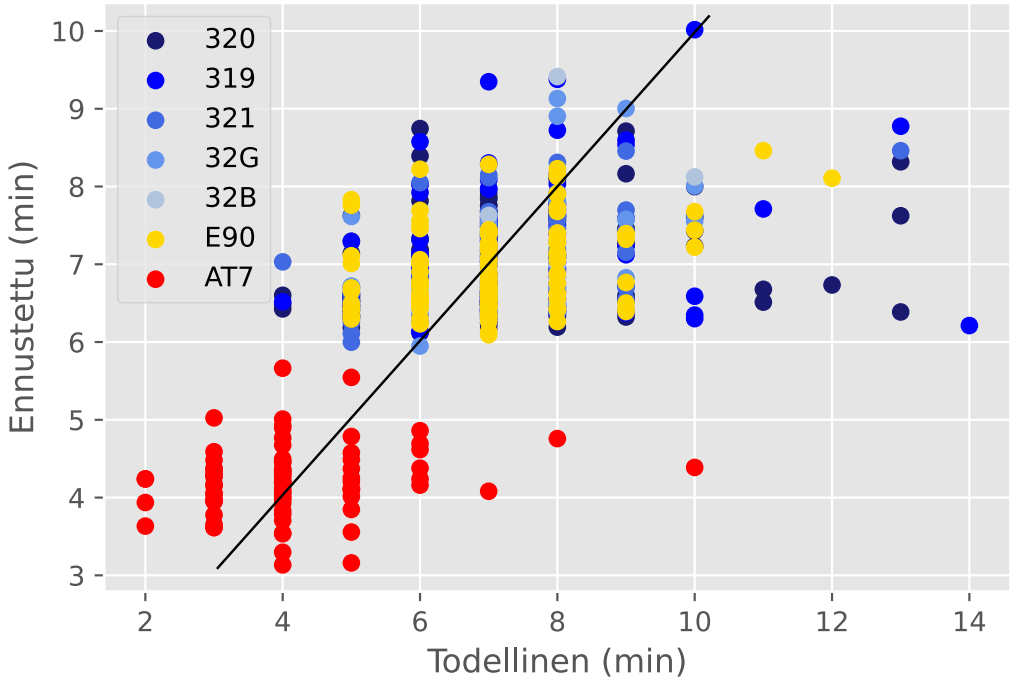
Kuviossa 34 on kuvattu testidatalla Helsingistä lähtevän liikenteen rullausaika ja ennuste. Hajontaa on enemmän kuin opetusdatalla, mutta keskimääräinen poikkeama pysyy hyvänä. Rullausaika Helsingin portilta lähtökiitotielle on keskimäärin 2,26 minuuttia, kun se opetusdatalla oli 1,37 minuuttia. Korrelaatio on 0,646, joka on kohtalainen, mutta hajonta on selvästi isompi kuin opetusdatalla. Opetusdatan korrelaatio oli 0,911. Isompi hajonta osoittaa, että kaikki poikkeamat eivät ole olleet ennustettavissa.



Kuvio 34. Helsingistä lähtevän liikenteen rullausaika ja ennuste testidatalla, $\rho_{X,Y} = 0{,}646$

Kuviossa 35 on kuvattu testidatalla Oulusta lähtevän liikenteen rullausaika ja ennuste. Korrelaatio on kohtalainen 0,620, joten hajontaa on enemmän kuin opetusdatalla. Opetusdataa vastaava korrelaation puute todellisen ja ennustetun ajan välillä toistuu, kuten tapahtui ope-

tusdatalla. Tämä vahvistaa oletusta, että Oulun lähtevän liikenteen rullausaikojen ennustetta pitää parantaa. Helsingin testidata antoi hieman vähemmän hajontaa kuin Oulun.

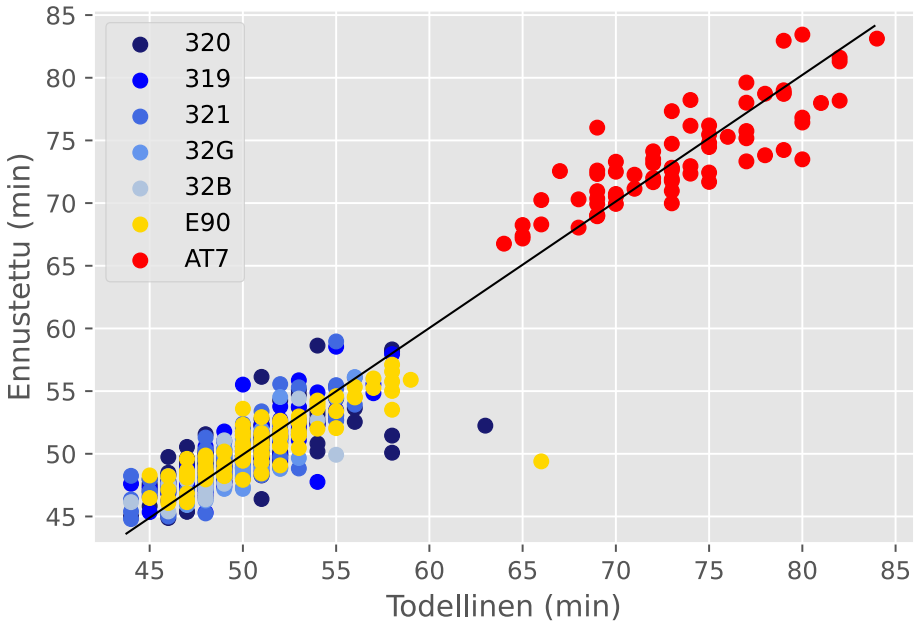


Kuvio 35. Oulusta lähtevän liikenteen rullausaika ja ennuste testidatalla, $\rho_{X,Y} = 0{,}620$

Rullausajan ennusteen keskimääräinen virhe Oulusta lähtevälle liikenteelle oli 0,91 minuuttia, kun vastaava luku oli opetusdatalla vain 0,80 minuuttia. Osaa virheestä selittänee todellisten aikojen ilmoittaminen kokonaisina minuutteina, mikä voi johtaa isoihin suhteellisiin virheisiin näin lyhyissä aikajaksoissa.

Kiitotielle rullauksen jälkeen tulee reittilennon vaihe, jonka ennusteet olivat erittäin vahvasti korreloivia opetusdatalla (0,99). Kokonaisuutta ajatellen reittilennon osuus ajallisesti on iso, joten tämä on tärkeä vaihe kokonaisuudelle. Kuviossa 36 on kuvattu testidatalla Helsinki–Oulu lentojen lentoaikaa ja ennustetta. Hajontaa on jälleen hieman enemmän kuin opetusdatalla. Lentoajan ennusteen keskimääräinen virhe Helsinki–Oulu välillä oli 1,34 minuuttia, kun se opetusdatalla oli 0,88 minuuttia. Testidatalla korrelaatio on erittäin vahva 0,973.

Kuviossa 37 on kuvattu testidatalla Oulu–Helsinki lentojen lentoaikaa ja ennustetta ilman kääntöaikaa. Hajontaa on hieman enemmän kuin opetusdatalla. Korrelaatio heikkeni opetuksesta testaukseen 0,992:sta 0,953:een, mikä on vielä erittäin hyvä tulos. Ilmeisesti Hel-

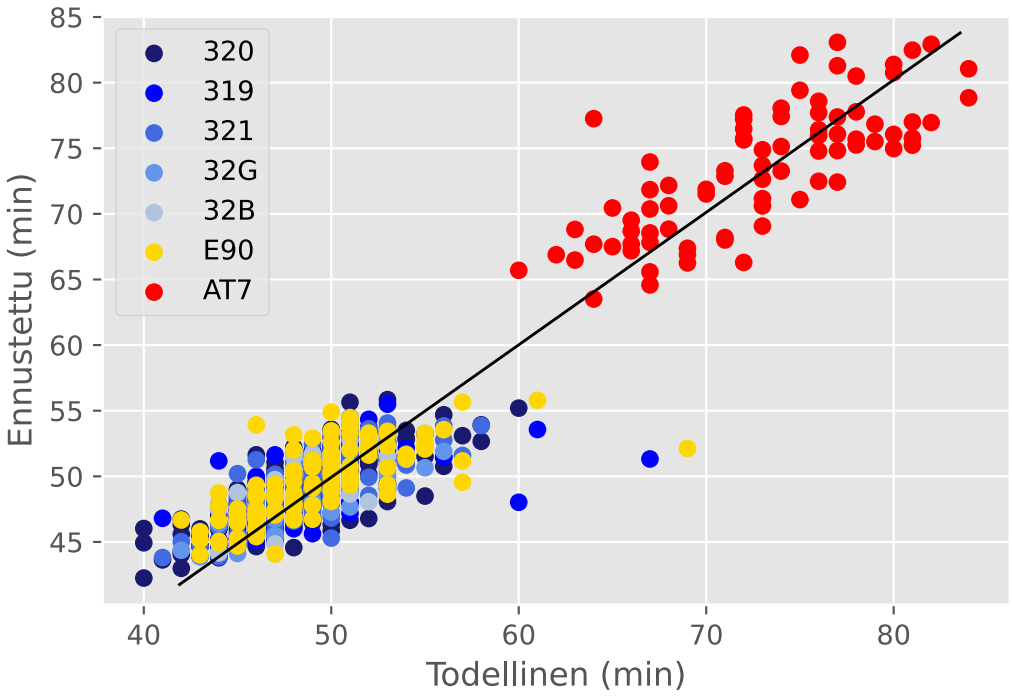


Kuvio 36. Helsinki–Oulu lentoajat ja ennuste testidatalla, $\rho_{X,Y} = 0{,}973$

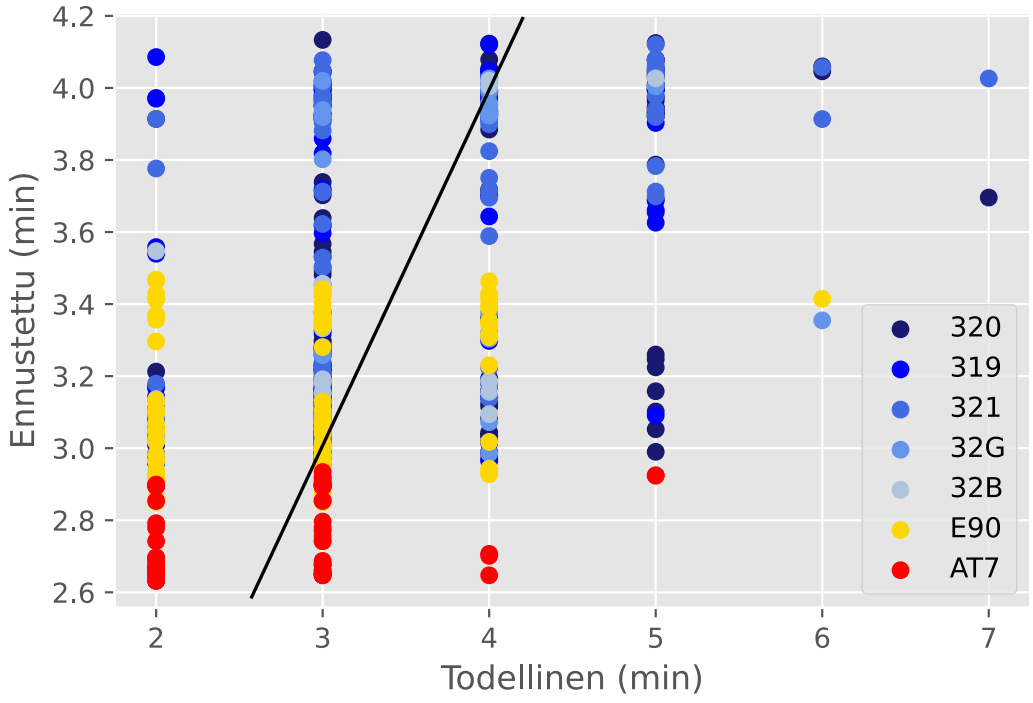
singin lähestymisten poikkeamat eivät ole täysin ennustettavissa. Lentoajan ennusteen keskimääräinen virhe Oulu–Helsinki välillä oli 1,89 minuuttia, kun se opetusdatalla oli 0,86 minuuttia.

Kuviossa 38 on kuvattu testidatalla Ouluun saapuvan liikenteen rullausaika ja ennuste. Korrelaatio oli ennusteista heikoin eli 0,548. Tämän vaiheen ennusteessa on selvästi parannettavaa. Myös aiemmin Oulun rullausajan ennustamisessa oli vaikeuksia, mutta aineiston puutteellisuutta ei saatu korjattua yrityksistä huolimatta. Rullausajan ennusteen keskimääräinen virhe Oulun tulevalla liikenteellä on 0,53 minuuttia, joka on sama kuin opetusdatalla. Ylioppimista ei tapahtunut, mutta se voi johtua oppimisen vaikeuksista.

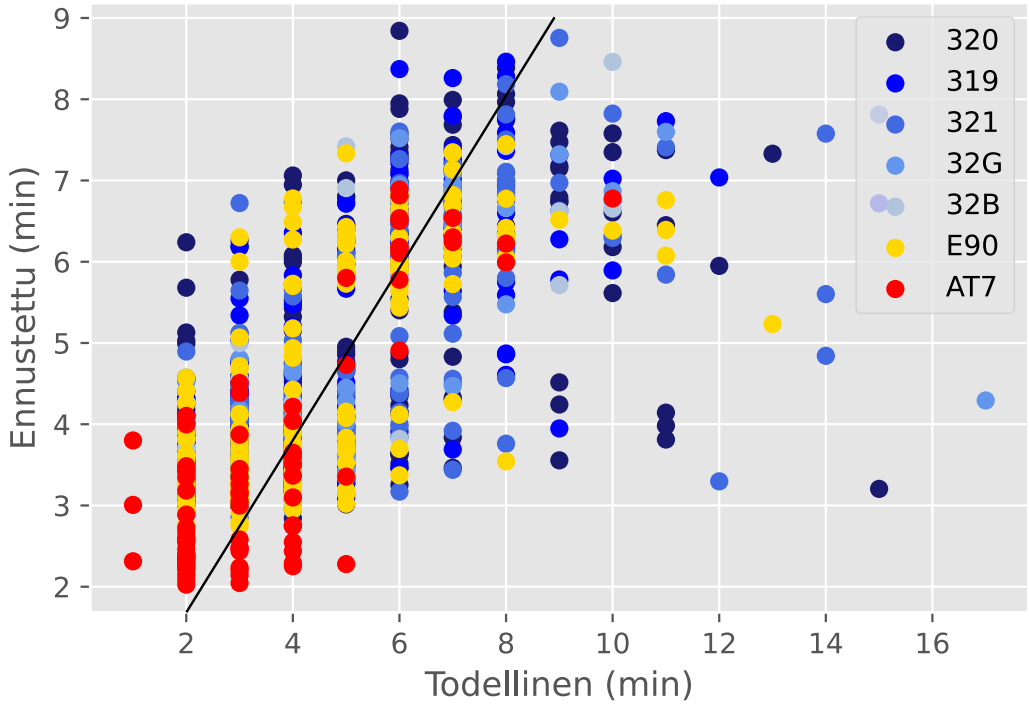
Helsinkiin saapuvan liikenteen rullaus on monimutkaisempi kuin Oulun, mutta tulos on johdonmukainen ja toiseen iteraatioon oli vielä käytettävissä kolme uuden tietämyksen kautta saatua syötemuuttujaa. Korrelaatio on 0,678. Ensimmäinen ennuste on kohtalainen testidatalla. Rullausajan ennusteen keskimääräinen virhe Helsinkiin tulevalla liikenteellä on 1,12 minuuttia, kun se opetusdatalla oli 0,82 minuuttia. Kuviossa 39 on kuvattu testidatalla Helsinkiin saapuvan liikenteen rullausaika ja ennuste.



Kuvio 37. Oulu–Helsinki lentoajat ja ennuste testidatalla, $\rho_{X,Y} = 0{,}953$



Kuvio 38. Ouluun saapuvan liikenteen rullausaika ja ennuste testidatalla, $\rho_{X,Y} = 0{,}548$

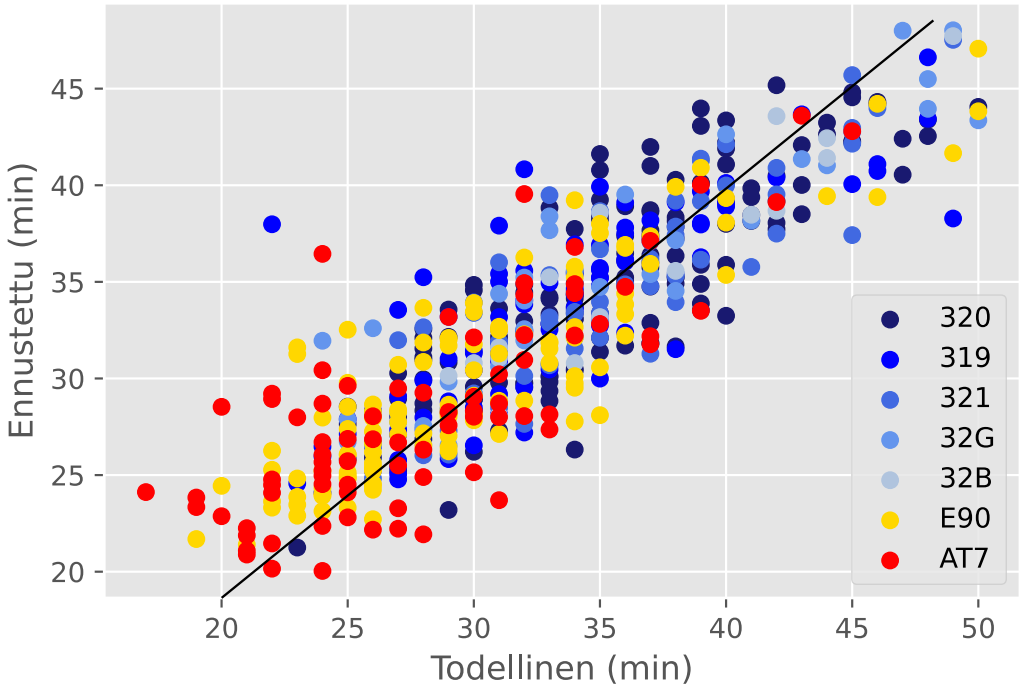


Kuvio 39. Helsinkiin saapuvan liikenteen rullausaika ja ennuste testidatalla, $\rho_{X,Y} = 0{,}678$

Lopuksi testataan lentoja yhdistävän Oulun kääntöajan ennuste testidatalla. Tässäkin vaiheessa hajontaa on odotetusti hieman enemmän kuin opetusdatalla. Korrelaatio on 0,880. Kuviossa 40 on kuvattu testidatalla kääntöaikaa ja sen ennustetta Oulussa. Opetusdatan yhteydessä mainitut toisen iteraation parannukset puuttuvat tästä testidatan vahvistamisesta, joten tätä ennustetta voidaan vielä parantaa. Kääntöajan keskimääräinen virhe Oulussa on 2,40 minuuttia, kun se oli opetusdatalla 1,54 minuuttia. Helsingin kääntöaikoja ei ennustettu, kuten aiemmin todettiin.

Kun testidatan mukaiset vaiheet kootaan yhteen, saadaan kuvaaja 41, jossa on 420 lentoparia Helsingistä Ouluun ja takaisin. Käytettävä aineisto on ennustusmenetelmälle täysin uusi.

Opetetuilla menetelmillä testidataa käyttäen meno-paluulennon ajan ja ennusteen tarkkuus paluulennon ajalle on keskimäärin neljä minuuttia viisi sekuntia. Tämä täyttää asetetun viiden minuutin tarkkuuden vaatimuksen. Ensimmäisen iteraation tuloksena voidaan todeta, että meno-paluulento on mahdollista ennustaa koneoppimisen menetelmin vaaditulla tarkkuudella. Korrelaatio on testidatan kokonaisennusteessa paluuajalle 0,927. Tulosta voi kuvata

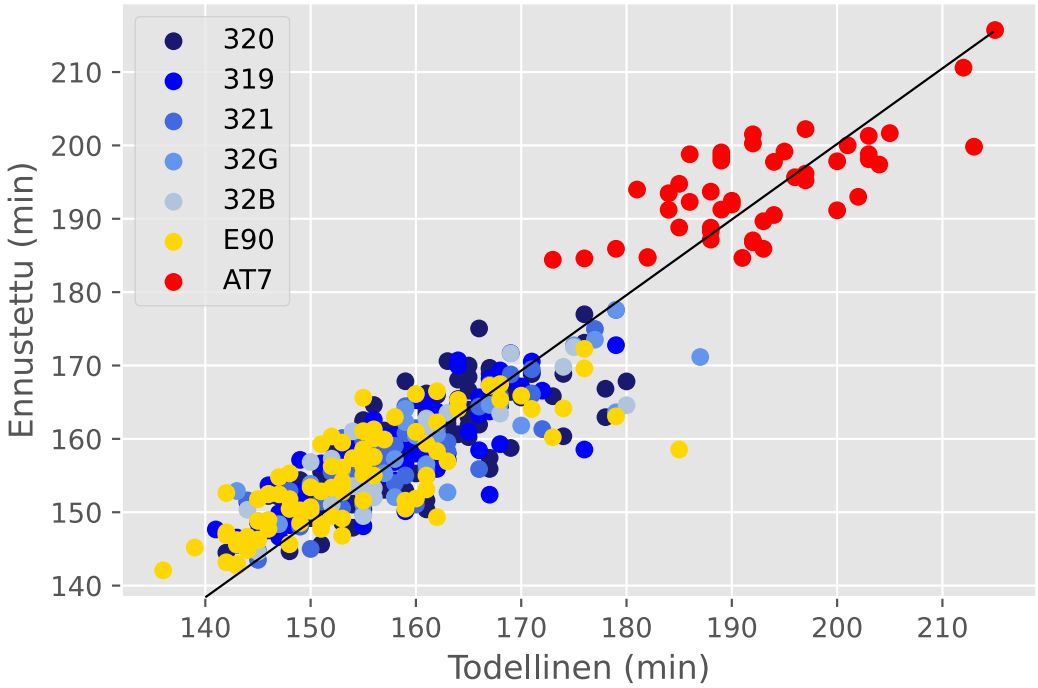


Kuvio 40. Kääntöajan ennuste Oulussa testidatalla, $\rho_{X,Y}$= 0,880

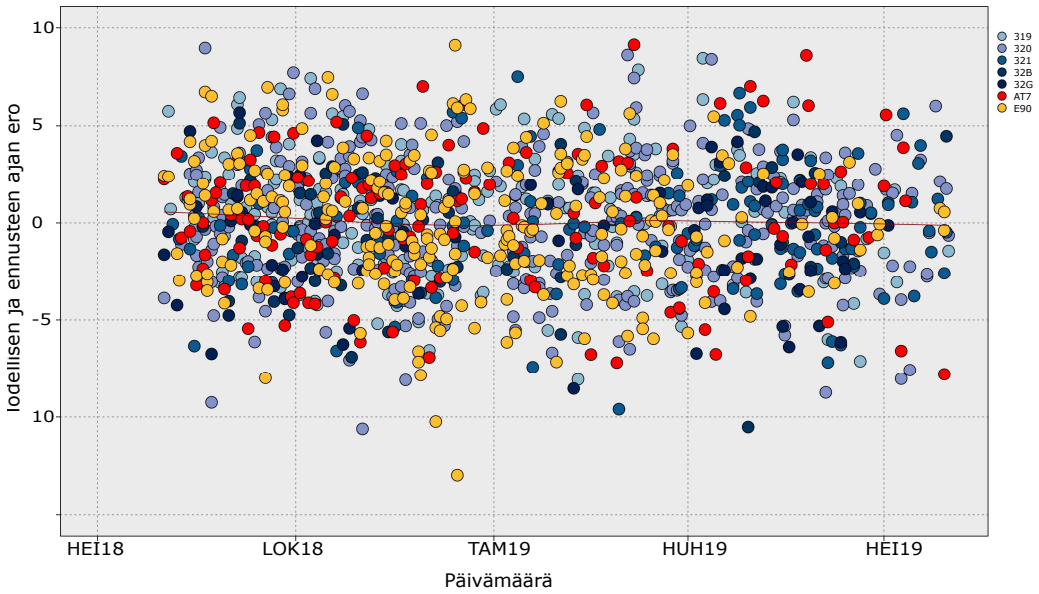
erittäin vahvaksi korrelaatioksi.

Lopuksi tarkastellaan kuvaajaa 42, jossa vaaka-akselilla on kalenteriaika ja pystyakselilla ennusteen poikkeama toteutuneesta ajasta. Poikkeama saadaan vähentämällä todellinen aika ennustetusta ajasta eli eron negatiivinen arvo tarkoittaa, että todellinen aika oli ennustettua pidempi. Poikkeamille on tyypillistä, että lento viivästyy poikkeamien vuoksi. Ilmeisesti poikkeamia pystytään ennakoimaan, koska ennusteen hajonta on tasainen ja korrelaatio on erittäin hyvä. Kuvaajan perusteella vuodenajoittain ennusteen erot jakautuvat tasaisesti. Konetyypeittäin ennusteen erot jakautuvat tasaisesti sekä poikkeaman määrän että kalenteriajan suhteen.

Taulukossa 7 on koottuna opetus- ja testidatan ennusteiden tarkkuudet vaiheittain ja kokonaisuutena. Mikäli vaiheen ajallinen pituus on pieni, ennustustarkkuuteen vaikuttavat todellisten aikojen suhteellisen suuret epätarkkuudet, koska aika on aina pyöristetty lähimpään minuuttiin. Käytännössä tuo pyöristyksen virhe on seuraavassa vaiheessa yleensä eri merkkinen. Näin meno-paluulennon kokonaisajan ollessa yli tunti, keskimääräinen pyöristysvirhe 15 sekuntia on selvästi alle prosentti lopputuloksessa. Kolmen minuutin rullauksessa suhteellinen



Kuvio 41. Helsinki–Oulu–Helsinki lentojen aika ja ennuste testidatalla (N = 420 paria), $\rho_{X,Y} = 0{,}927$



Kuvio 42. Helsinki–Oulu–Helsinki lentojen aika ja ennusteen virhe testidatalla

osuus on paljon suurempi. Esimerkiksi Oulussa rullaus portille ajassa on 0,87 minuuttia keskimääräinen ennusteen poikkeama, mutta todellisissa rullausajoissa on virhe 0,25 minuuttia.

Lisäksi todellisten aikojen virhe vaikeuttaa mallin opetuksen onnistumista.

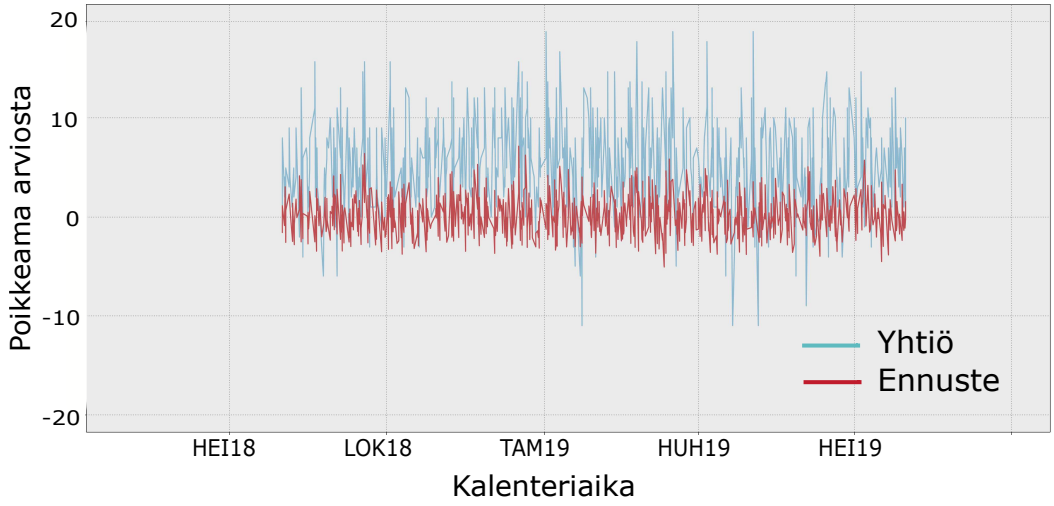
Taulukossa ovat absoluuttiset poikkeamat todellisesta ajasta minuutteina. Lisäksi testidatalle ovat Pearsonin korrelaatiot. Taulukossa ei ole eroteltu konetyyppejä. Huonoin korrelaatio on molemmissa Oulun rullausajoissa testidatalla, mihin myös vaikuttaa lyhyet keskimääräiset rullausajat 3,4 minuuttia portille ja 6,5 minuuttia portilta. Lyhyillä aikaväleillä korrelaatiota huonontaa todellisen ajan pyöristys lähimpään minuuttiin. Oulun lentoaseman aineiston kattavuutta ei saatu Helsinkiä vastaavaksi ja merkittäviä syötemuuttujia jäi pois laskennasta.

Vaihe	Kesto (min)	Opetus-virhe (min)	Opetus $\rho_{X,Y}$	Testi-virhe (min)	Testi $\rho_{X,Y}$
Rullaus kiitotielle HEL	10,5	1,37	0,911	2,26	0,646
Reittilento HEL–OUL	52,5	0,88	0,991	1,34	0,973
Rullaus portille OUL	3,4	0,87	0,585	0,53	0,548
Kääntöaika OUL	32,6	1,54	0,953	2,40	0,880
Rullaus kiitotielle OUL	6,5	0,80	0,753	0,91	0,620
Reittilento OUL–HEL	51,6	0,86	0,992	1,89	0,953
Rullaus portille HEL	4,6	0,82	0,880	1,12	0,678
HEL–HEL	161,2	2,50	0,974	4,10	0,927

Taulukko 7. Ennusteiden tarkkuudet ja korrelaatiot (MAE ja Pearsonin korrelaatio)

Edellinen laskenta koski koko meno-paluulennon ennustetta. Ennusteen toinen mahdollinen käyttötapa on ennustaa aikataulua lennon vaiheiden edetessä. Finnair ilmoittaa lentoonlähdön jälkeen arvioidun saapumisajan määräkentän tuloportille. Kuvaajassa 43 on kuvattu sinisellä Helsinkiin tulevien lentojen poikkeaminen arvioidusta tuloajasta portille. Negatiivinen poikkeama tarkoittaa, että lento tulikin ilmoitettua aiemmin portille. Punaisella on kuvattu vastaavasti tutkimuksen ennustaman tuloajan poikkeaminen oikeasta tuloajasta. Finnairin arvion virhe on keskimäärin 4,7 minuuttia ja tutkimuksen arvion virhe on keskimäärin 1,2 minuuttia. Tutkimus ennustaa neljä kertaa tarkemman arvion tuloajalle.

Lähdekirjallisuudessa ennustettiin pääasiassa myöhästymisiä eri vaiheissa, mutta kolmessa



Kuvio 43. Oulu–Helsinki lennon arvioidun tuloajan poikkeaminen todellisesta ajasta

tutkimuksessa löytyi aika-arviot tietyille vaiheille. Kyseiset tutkimukset olivat eri lentokentiltä kuin tässä tutkimuksessa, joten niiden vertailu ei ole täysin luotettavaa, mutta antaa suuntaa tulosten onnistumisesta.

Lähdekirjallisuudessa Hassel (2019) ennusti kääntöaikaa ja saavutti noin 4,3 minuutin absoluuttisen virheen. Hänen syötemuuttujissa olivat monipuoliset tiedot rahdista ja matkustajista. Vastaava absoluuttinen virhe tämän tutkimuksen kääntöajoille oli testidatalla 2,4 minuuttia. Halmesaari (2020) puolestaan ennusti kääntöajan 2,8 minuutin virheellä ja hänellä oli myös yksityiskohtaiset rahti ja matkustajatiedot lennoille.

Kiitotielle kuluvaa rullausaikaa ennusti Lee ym. (2015b) ja saavutti 4-5 minuutin absoluuttisen virheen, kun tässä tutkimuksessa virheet olivat Helsingissä 2,3 ja Oulussa 0,9 minuuttia.

Tutkimuksen ensimmäisen iteraation kaikki ennusteet ja tulokset ovat koottu yhteen ja voidaan arvioida, kannattaako jatkaa seuraavaan kierrokseen. Tässä vaiheessa jo saadaan viitteitä, kunka hyvin koneoppiminen selviytyy tavoitteesta, kun tulosta vielä parannetaan saadun uuden tietämyksen avulla.

6 Johtopäätökset, pohdinta ja jatkotutkimuksen aiheet

Ensimmäinen iteraatio täytti tutkimuksen tavoitteen tietyin ehdoin. Oletuksena oli, että ulkoiset syöte-ennusteet ovat tarkkoja. Tässä luvussa pohditaan tulosten luotettavuutta ja erityispiirteitä. Tietämyksen lisääntyminen ensimmäisessä iteraatiossa lisääntyi suunnittelutieteen mukaisesti. Uusi tietämys antaa aihetta pohtia toisen iteraation kehityksen suuntaa ja tavoitetta. Kuten usein tavoitteen asettelussa on todettu, asiakkaan ongelman ratkaisu takaa hyvän tavoitteen ja merkityksellisen tutkimuksen. KM-mallissakin viimeisenä vaiheena on hyödyn saaminen tuloksista (Äyrämö 2006).

6.1 Johtopäätökset

Tutkimuksen tavoitteena oli selvittää, onko tekoälyllä mahdollista ennustaa meno-paluulennon aikataulu vuorokautta ennen lennon lähtöä viiden minuutin tarkkuudella. Tutkimukselle asetetut tavoitteet vahvistettiin ja tuloksena oli neljän minuuttia viisi sekuntia. Tutkimusmenetelmänä käytettiin suunnittelutiedettä ja ennustamisessa koneoppimisen menetelmänä eXtreme Gradient Boosting -menetelmän ennustetta.

Ongelma jaettiin meno-paluulennon vaiheisiin ja jokainen vaihe ennustettiin erikseen. Vaiheiden ennusteet summattiin kokonaisennusteeksi. Ensimmäisessä iteraatiossa käytettiin syötemuuttujissa oikeaa ajantasaista tietoa ja ulkoisten syöte-ennusteiden vaikutusta tutkitaan myöhemmin, mikäli tarkennettu tavoite sitä edellyttää. Vaiheistus ja tarkat syötemuuttujat selkiyttivät ongelmien ratkaisua ja mahdollistivat vaiheiden optimoinnin ilman ulkoisten syöte-ennusteiden epätarkkuuden vaikutusta mallin opetukseen. Jokaista vaihetta optimoitiin, kunnes saavutettiin riittävä tarkkuus. Suunnittelutieteen mukaisesti syötteitä rikastettiin tietämyksen lisääntyessä. Ylioppimista pyrittiin estämään oppimisessa käytetyillä menetelmillä. Tuloksen vahvistamiseen käytettiin erillistä testidataa, jolla lopuksi testattiin ennustaminen ennalta tuntemattomalla aineistolla. Ylioppimista ei havaittu ja lopputulos oli testidatalla korrelaation perusteella erittäin tarkka. Myös vakiintuneeseen paluulennon tuloarvioon saatiin neljä kertaa tarkempi arvio tutkimuksen ennusteella.

Kehitettyä mallia kokonaisennusteelle ei oltu aiemmin tutkittu. Aiempia tutkimuksia hyö-

dynnettiin koneoppimismenetelmän ja syötemuuttujien valinnoissa. Tutkimuksen ennusteiden tuloksia ei täysin voida vertailla aiempiin tutkimuksiin, koska tässä tutkimuksessa ei tehty ennusteen ja todellisen aikataulun vertailua, joka kertoisi mahdolliset myöhästymistapaukset. Kyseinen tieto on saatavilla, mutta ei kuulu tutkimuksen tavoitteisiin. Mikäli tietoa tarvitaan jatkotutkimuksissa, tehty laskentamalli on helppo muuttaa tarpeen mukaiseksi. Joiltain osin tehtiin vertailua ja tämä tutkimus ennusti paremmin, mutta tutkimusaineistot eivät ole vertailukelpoisia.

Saatu tietämys aiheesta mahdollistaa erityyppisten ratkaisujen mallintamisen eri tarpeisiin, mistä kerrotaan seuraavassa pohdinnassa. Tietämyksen merkitys lopputulokseen oli odotettavissa aiempien teorioiden perusteella.

6.2 Pohdinta

Mallin rakentaminen eri vaiheista osoittautui hyväksi ratkaisuksi. Näin oli mahdollista optimoida kutakin vaihetta, kunnes saavutettiin riittävä tarkkuus. Vaiheiden vaatimat syötemuuttujat osoittautuivat niin erilaisiksi, että yhden mallin ratkaisuun tarvittaisiin kymmeniä syötemuuttujia ja mallin moniulotteisuus ei olisi lupaava vaihtoehto lopulliseksi ratkaisuksi. Monen vaiheen malli mahdollistaa lopullisen sovelluksen tekemisen niin, että lennon edetessä tai lähtöhetken lähestyessä, voidaan lopputulosta tarkentaa. Lähtöhetken syötteet tarkentuvat ennusteajan lyhentyessä, joten ennuste muuttuu käytännössä kolmen tunnin välein. Lisäksi on mahdollista muuttaa konetyyppien ja lähtöporttien syötemuuttujia ja vaikuttaa ennusteen lopputulokseen. Jakamalla lennon ennusteet neljään vaiheeseen voidaan helpommin lisätä ennustuksiin uusia lentokenttiä. Uudelta yhteydeltä tarvittaisiin vain reittilennon ennuste.

Ennusteen jakamista voisi lisätä jakamalla vaiheet konetyypeittäin. Jakamisella varmistettaisiin, ettei eri konetyyppien ominaisuudet häiritse toistensa ennusteita. Samalla saataisiin helpommin eriteltyä, jos ongelmia ilmenee vain jonkin konetyypin ennusteissa. Kuvaajissa ennusteet ovat eriteltyinä konetyypeittäin, joten tyyppikohtaiset ongelmat todennäköisesti erottuisivat kuvaajissa. Toinen vaihtoehto olisi yhdistää ennusteita, jotta välivaiheet sisältyisivät yhteen ennusteeseen, joka yhdistää kaikkien vaiheiden syötemuuttujat. Näin tehtäisiin yksi ennuste koko meno-paluulennolle. Kolmas vaihtoehto olisi tutkia, millaisen tarkkuus on

mallilla, jossa on kukin vaihe ennustettu vain 2–3 tärkeimmällä syötemuuttujalla.

Ytimenä olevan kokonaisennusteen syöte-ennusteet korvattiin oikeilla tiedoilla ja näin saatiin tieto optimimallin kyvystä. Näin pystyttiin ydinennusteen mahdollisimman hyvään opetukseen syöte-ennusteiden epätarkkuuksista huolimatta. Ulkoiset syöte-ennusteetkin voidaan valita ja rikastaa kehityksen seuraavassa iteraatiossa, kun ydinennuste on tarkka ja luotettava. Ulkoisista syöte-ennusteista haastavin on kiitotie-ennuste, mutta siihen on hyvät perusteet tässä tutkimuksessa. TAF-sanomien käyttö mallin syötemuuttujana on mielenkiintoinen ja lisää varmasti tulevissa tutkimuksissa tietämystä sopivasta sääennusteen käytöstä.

Eri vaiheissa tuli uutta tietämystä koneoppimisen kyvyistä tietyissä toimissa. Esimerkiksi lentoaikaa voidaan ennustaa tarkasti vähäisillä syötemäärillä. Ylätuuliennusteiden tarkkuus mahdollistanee niin hyvän pitkänajan ennusteen, että ennakkotietoa liikenteen poikkeamista saadaan pitkilläkin aikaväleillä.

Ennuste koko meno-paluulennolle on saatavilla tarkemmin ja helpommin kuin lennonsuunnitteluohjelmilla, koska koneoppiminen löytää oikeat muuttujat ja hyödyntää tietoa toistuvista poikkeamista. Saatu perustieto syötemuuttujien merkityksellisyyksistä mahdollistaa eri kehitysmallien arvioinnin. Samalla voidaan pyrkiä tarkentamaan piirteiden valintaa paremmilla piirteiden valintaan keskittyvien mallien käytöllä. Uuden tietämyksen mukaan voidaan kehittää mallia useiden eri tarpeiden mukaisesti. Mallia voidaan vielä tarkentaa uuden tietämyksen mahdollistamilla tarkennuksilla syötemuuttujiin, jolloin mallin tarkkuus säilyy, vaikka ennusteaikaa pidennetään.

Suunnittelutieteen käyttö lisäsi perustietoa aiheen käsittelystä ja mahdollistaa kehityksen tiettyjen tarpeiden mukaisesti. Tietämyksen tärkeys ja tiedon esikäsittelyn vaatima aika osoittautui aiempien teorioiden mukaisiksi. Tietämys oli tärkein osa tulosta. Syötemuuttujia ei olisi saatu hyviksi ilman tietämystä. Nämä kokemukset vahvistuivat ja tiedossa olevat tosiasiat tulivat uudelleen varmistettua. Aiempi käsitykseni aihepiirin asiantuntijoiden merkityksestä vahvistui. Poikkeuksellisesti tällä kertaa aihepiirin asiantuntijalla oli myös ymmärrys tekoälymenetelmän vaatimuksista. Koneoppimisen optimointi hyvin rikastetulla aineistolla voidaan tehdä lopuksi vakiintunein menetelmin.

Suunnittelutieteen mukaisesti, sovelluksen julkaisu vaatii uuden iteroinnin tarkennetun ta-

voitteen mukaisesti. Erilaiset tarpeet ennusteesta voidaan huomioida seuraavassa iteraatiossa. Jo havaitut piirteet voidaan kehittää paremmiksi sen mukaan, mitä tavoitellaan.

Tavoitteena voi olla esimerkiksi pidemmät ennusteajat, jolloin saadaan riittävä tarkkuus hyvissä ajoin saatavalle varoitukselle tulevien poikkeamien mahdollisuudesta. Tällaiseen malliin voidaan keventää syötemuuttujien määrää, tinkiä tarkkuudesta ja optimoida syötteitä. Mikäli halutaan vuorokauden ennakkoon tarkkaa tietoa poikkeamista, tulee ulkoisia syöteennusteita tutkia ja hakea malli, joka toimii parhaiten vuorokauden ennusteella. Tällöin tulee painottaa TAF-ennusteita ja käytettävän kiitotien ennustetta. Joillekin paras vaihtoehto on päivän aikana tehty ennuste, jossa tarkkuutta voidaan entisestään parantaa ja valita syötemuuttujat, joilla parhaiten voidaan vaikuttaa toteutuvaan ennusteeseen. Joillekin riittää, että päivän aikana olevat tapahtumat ennustetaan tarkasti ilman puuttumista tilanteen kehitykseen. Näissä tapauksessa keskitytään METAR-sääsanomiin ja lennon vaiheiden edetessä tarkennettavaan ennusteeseen.

Vaikka nyt keskityttiin koneoppimisella ennustamiseen, saadulla tietämyksellä on mahdollista kehittää myös sääntöpohjainen laskenta. Laskennassa käytettäisiin merkittäviksi todettuja syötemuuttujia. Mielenkiintoista oli, että lentovaiheen ennuste on mahdollista tehdä hyvin vähäisillä syötemuuttujilla ja tarkkuus voi olla parempi kuin lennonsuunnitteluohjelmien laskelmat. Kun tähän lisätään rullausaikojen ennusteet, kokonaisuuden tarkkuudet ovat lennonsuunnitteluohjelmistoja paremmat. Koneoppimisen luonteesta johtuen nämä ennusteet eivät kuitenkaan täytä operatiivisia viranomaisvaatimuksia, mutta ovat hyödyllisiä apuvälineitä lentotoiminnan ennakoinnille ja suunnittelulle. Voisi olla hyödyllistä laskea esimerkiksi polttoainekulutukset molemmilla tavoilla ja tankata lentokone niin, ettei kummankaan laskelman arviota aliteta. Näin tulisi huomioitua paremmin tavanomaiset poikkeamat kyseiselle lennolle.

Kaikissa vaihtoehdoissa lentokenttäverkoston lisäämisen helppoutta tulisi painottaa mallin kehittämisessä. Aluksi olisi helppoa lisätä lentokenttä-verkkoon pohjoiseen suuntautuvaa liikennettä. Rovaniemen aineisto on jo valmiiksi rikastettu. Lyhyenä reittinä Tukholman suunta olisi hyödyllinen ja mielenkiintoinen samanlaisena kenttänä kuin Helsinki–Vantaa.

Joitain tärkeitä tietoja puuttui Finnairin lentotiedoista, kuten matkustajamäärä ja rahdin mää-

rä. Ne voitaisiin korvata seuraavassa iteraatiossa koneen maksimipaikoilla ja maksimi lentoonlähtömassalla. Koska konetyyppi optimoidaan parhaalle täyttöasteelle eli matkustajamäärälle, paikkamäärä voisi korreloida täyttöasteen kautta matkustajien määrää. Tällä tavoin tutkimus säilyttää julkisen statuksen ja voitaisiin päästä vielä tarkempiin kääntöajan ennusteisiin. Lennon numeroakin voisi käyttää ennusteessa kääntöaikoihin, jos sama lento on aina täynnä samana viikonpäivänä tai kellonaikana.

Eräs kääntöaikoihin liittyvä oleellinen ennuste jäi vielä puuttumaan – aamun lähtövarmuus. Sillä ennustettaisiin aamun ensimmäisen lähdön myöhästymistä. Se poikkeaa vaiheiden ennusteista, koska siihen vaikuttaisi myös lentokentän toiminnot sekä matkustajien sekä miehistön siirtyminen lentokentälle ja kentällä. Ennuste olisi sikäli tärkeä, että aamun myöhästyminen vaikuttaa koko päivän aikatauluihin.

Muutakin tietoa jäi vielä toiseen iteraatioon, mutta mallin muokkausta ei kannattane jatkaa ennen uuden tavoitteen asettamista. Samoja syötemuuttujia käyttäen aineiston kokoa voitaisiin vielä lisätä huomattavasti lisäämällä lentojen määrää eri vuosilta ja eri kenttäväleiltä. Tämä tapa ei vaatisi paljoa lisätyötä, mutta parantaisi varmasti tarkkuutta. Paras tapa lisätä ennusteiden tarkkuutta on aineiston koon kasvattaminen. Näin pystyttäisiin vielä paremmin ennustamaan säännönmukaisia poikkeamia liikenteestä.

Mallilla voidaan ennustaa aikataulua alkaen kustakin lennon vaiheesta. Tarkimmillaan ennusteet ovat, kun ennusteaika on lyhyimmillään. Tavoitteena ollut viiden minuutin tarkkuus tuloportille täyttyy juuri lentoyhtiön arvioissa paluulennon lentoonlähdön jälkeen tehtynä, joka ennusteaikana tarkoittaa yksi tai kaksi tuntia etukäteen ennustamista. Tutkimuksen ennuste on siinä vaiheessa neljä kertaa tarkempi. Tutkimuksen ennusteella saadaan parempia tarkkuuksia jo koneen aloittaessa lähdön portilta meno-paluulennolle. Mikäli syötemuuttujina käytetyt ennusteet pitävät paikkansa, lentoyhtiön arvioita tarkempi ennuste saadaan ennustettua jo edellisenä päivänä. Sikäli ennusteen käytölle on kiistattomat perusteet.

Lopulta tutkimuksen tulos oli odotusten mukainen ja varmisti oman ennakkokäsitykseni tekoälyn ja lentoliikenteen yhdistämisen mahdollisuuksista ja hyödyistä. Yllätyksenä ja uutena asiana tutkimuksessa huomasin syötemuuttujien rikastamisen tärkeyden. Vaikkakin se olisi pitänyt aiempien tutkimusten mukaisesti olla odotettavissa. Tietämyksen lisääntyessä uusien

mahdollisuuksien ja ratkaisujen näkeminen kirkastui itselleni ja toivottavasti kirkastuu myös monelle lukijalle.

6.3 Jatkotutkimuksen aiheita

Jatkotutkimuksen aiheet määräytyvät tarkemmista tavoitteista, mutta mahdollisuudet kehittämiseen on jo huomioitu vaiheiden analyyseissä. Ennustetta voitaisiin tehdä pidemmälle aikavälille ja yleisemmillä syötemuuttujilla. Toisaalta tarpeet voivat olla samalle päivälle vielä tarkentaen syötemuuttujia ja vaiheiden ennusteita. Vaativin välimalli olisi vuorokauden ennuste, jossa ulkoisten syöte-ennusteiden epätarkkuutta korvattaisiin uusilla syötemuuttujilla ja optimoimalla epätarkkojen vaiheiden ennusteita. Samalla huomioitaisiin syötemuuttujien vaikutus ennusteeseen. Esimerkiksi konetyypin, lähtöportin tai tuloportin vaihtaminen vaikuttaisi ennusteeseen.

Kaikki mallit vaatisivat lentokenttäverkon laajentamisen ja siihen olisi jo valmiina Rovaniemen rikastetut tiedot. Rovaniemen jälkeen kannattaa jatkaa Helsinki–Tukholma välin malliin, koska näin saataisiin kaksi erityyppistä kenttäväliä ja uutta tietämystä mallin jatkokehitykseen edelleen. Kiitotie-ennuste liittyisi myös kaikkiin vaihtoehtoihin. Myös jo valmiit mallit voitaisiin opettaa uudelleen esimerkiksi nelinkertaisilla aineistomäärillä.

Joka tapauksessa seuraavassa tutkimuksessa voidaan keskittyä uuden tietämyksen mahdollistamiin syötemuuttujiin. Lisäksi voidaan painottaa heikkojen vaiheiden parannuksiin. Mallia voidaan painottaa pidemmän ajan ennusteisiin kevyemmillä syötemuuttujilla tai tarkempiin ennusteisiin kyseisen päivän ajalle.

Tämä tutkimus osoitti, että nämä vaihtoehdot ovat mahdollisia ja tietämys oikeanlaiseen kehitykseen on helppo tehdä. Myöhästymisten ennusteista voidaan siirtyä tekemään ennusteita toteutuvasta aikataulusta omiin tarpeisiin. Nyt on saatu lupaava tulos tutkimukselle ja iteraation toiselle kierrokselle tulisi asettaa uusi tavoite. Tavoitteeseen vaikuttaa kuka on asiakas ja mikä on heidän tarpeensa. Pääasialliset sidosryhmät, joille kehitys vaatisi omaa räätälöintiä olisivat Traficom, Fintraffic, Finavia, Finnair ja lentomatkustajat.

7 Yhteenveto

Tutkimus osoitti, että koneoppimisella voidaan tehdä erittäin tarkka aikatauluennuste tavoitteen mukaisesti. Meno-paluulennon keskimääräinen poikkeama todellisesta paluuajasta oli neljä minuuttia eli parempi kuin asetettu tavoite.

Vaiheittain ennustaminen oli hyödyksi mallin kehittämisessä sekä mahdollisuutena tarkentaa ennustetta vaiheittain lennon edetessä. Paluulennon alkaessa voitiin ennustaa portille tulo jopa yhden minuutin tarkkuudella, mikä oli neljä kertaa tarkempi kuin nykyiset lentoyhtiön arviot.

Saatu uusi tietämys vaiheiden ennustamisesta mahdollistaa erilaiset tavoitteet sekä uuden iteraation jälkeen ennusteen tarkkuuden parantamisen edelleen. Tälläkin iteraatiolla saadaan kokonaisennusteelle erittäin vahva korrelaatio ja mahdollisuus jopa tukea perinteisiä lennonsuunnittelun laskelmia tarkoilla ennusteilla. Lentoliikenteessä tapahtuvat poikkeamat voidaan näin sisällyttää lennonsuunnitteluun, jolloin ennuste tarkentaa laskettua suunnitelmaa.

Kun ennustetaan seuraavan päivän lentoja, saattavat ulkopuoliset syöte-ennusteet aiheuttaa epätarkkuutta lopputulokseen. Perusmenetelmä on kuitenkin tarkka ja sitä voidaan kehittää tarpeiden mukaisesti eri sidosryhmille. Mikäli vastaavia malleja halutaan käyttää lentoliikenteen hyödyksi, tutkimus osoitti, että tekoälyllä voidaan ennustaa aikataulupoikkeamia riittävän tarkasti.

Lähteet

Airline On-Time Statistics and Delay Causes. 2022. United States Department of Transportation. Viitattu 9. helmikuuta 2022. https://www.transtats.bts.gov/OT_Delay/ot_delaycause1.asp?6B2r=E&20=E.

Akoglu, Haldun. 2018. "User's guide to correlation coefficients". *Turkish Journal of Emergency Medicine* 18 (3): 92. ISSN: 2452-2473. https://doi.org/https://doi.org/10.1016/j.tjem.2018.08.001. https://www.sciencedirect.com/science/article/pii/S2452247318302164.

Baomar, Haitham, ja Peter Bentley. 2017. "Autonomous landing and go-around of airliners under severe weather conditions using Artificial Neural Networks", 162–167. Lokakuu. https://doi.org/10.1109/RED-UAS.2017.8101661.

Bell, Jason. 2020. *Machine learning : Hands-on for developers and technical professionals.* 2. painos. 82–88. ISBN: 978-1-119-64219-0.

Ben Ahmed, Mohamed, Farah Zeghal Mansour ja Mohamed Haouari. 2018. "Robust integrated maintenance aircraft routing and crew pairing". *Journal of Air Transport Management* 73:15–31. ISSN: 0969-6997. https://doi.org/https://doi.org/10.1016/j.jairtraman.2018.07.007.

Bergstra, James, ja Bengio Yoshua. 2012. "Random Search for Hyper-Parameter Optimization", ISSN: 1532-4435.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning.* 1. painos. Springer New York.

Breiman, Leo. 2001. "Random forests". *Machine learning* 45 (1): 5–32.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall ja Philip W. Kegelmeyer. 2002. "SMOTE: synthetic minority over-sampling technique". *Journal of artificial intelligence research* 16:321–357.

Chen, Jun, ja Meng Li. 2019. “Chained Predictions of Flight Delay Using Machine Learning”. Teoksessa *AIAA Scitech 2019 Forum.* https://doi.org/10.2514/6.2019-1661. https://arc.aiaa.org/doi/abs/10.2514/6.2019-1661.

Chen, Tianqi, ja Carlos Guestrin. 2016a. “XGBoost: A Scalable Tree Boosting System”, 785–794. Elokuu. https://doi.org/10.1145/2939672.2939785.

———. 2016b. “XGBoost: A Scalable Tree Boosting System”, 785–794. Elokuu. https://doi.org/10.1145/2939672.2939785.

Choi, Sun, Young Jin Kim, Simon Briceno ja Dimitri Mavris. 2016. “Prediction of weather-induced airline delays based on machine learning algorithms”, 1–6. Syyskuu. https://doi.org/10.1109/DASC.2016.7777956.

Codina, Ramon Dalmau, Seddik Belkoura, Herbert Naessens, Franck Ballerini ja Sebastian Wagnick. 2019. “Improving the predictability of take-off times with Machine Learning : a case study for the Maastricht upper area control centre area of responsibility”.

Cook, A. J., ja G. Tanner. 2015. “European airline delay cost reference values”. *Technical report,* 4–6. Viitattu 9. helmikuuta 2022. http://www.eurocontrol.int/publications/european-airline-delay-cost-reference-values.

Darlington, Richard, ja Andrew Hayes. 2017. “Regression analysis and linear models”. *New York, NY: Guilford,* 603–611.

Denning, Peter J. 1997. “A New Social Contract for Research”. *Commun. ACM* (New York, NY, USA) 40, numero 2 (helmikuu): 132–134. ISSN: 0001-0782. https://doi.org/10.1145/253671.253755. https://doi.org/10.1145/253671.253755.

Diepen, Guido, B. F. I. Pieters, J. M. Akker ja J. A. Hoogeveen. 2009. “Robust planning of airport platform buses”. *Computers & Operations Research* 40 (joulukuu).

Dunham, Margaret H. 2003. *Data mining introductory and advanced topics.* Upper Saddle River. ISBN: 0130888923 9780130888921.

Eurocontrol. 2020. “All-causes delay and cancellations to Air Transport in Europe for 2019”. *CODA Digest* (huhtikuu).

Fayyad, U. M., G. Piatetsky-Shapiro ja P. Smyth. 1996a. “From data mining to knowledge discovery: an overview, American Association for Artificial Intelligence”. *AI Magazine,* numero 11, 1–34.

———. 1996b. “From data mining to knowledge discovery: an overview, American Association for Artificial Intelligence”. *AI Magazine,* 10–11.

Gregorutti, Baptiste, Bertrand Michel ja Philippe Saint-Pierre. 2017. “Correlation and variable importance in random forests”. *Statistics and Computing* 27 (toukokuu). https://doi.org/10.1007/s11222-016-9646-1.

Guo, Zhen, Bin Yu, Mengyan Hao, Wensi Wang, Yu Jiang ja Fang Zong. 2021. “A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient”. *Aerospace Science and Technology* 116:106822. ISSN: 1270-9638. https://doi.org/https://doi.org/10.1016/j.ast.2021.106822.

Halmesaari, Eppu. 2020. “Interpretable machine learning for prediction of aircraft turnaround times”.

Hassel, O. F. J. Van. 2019. “Predicting the Turnaround Time of an Aircraft: A Process Structure Aware Approach”.

Haykin, Simon. 2009. *Neural Networks and Learning Machines.* 3. painos. ISBN: 978-0-13-147139-9.

Herbert, Simon A. 1996. *The sciences of the artificial.* MIT press.

Hevner, Alan R. 2007. “A Three Cycle View of Design Science Research”. 19 (2): 87–92. https://aisel.aisnet.org/sjis/vol19/iss2/4.

Hevner, Alan R., Salvatore T. March, Jinsoo Park ja Sudha Ram. 2004. “Design Science in Information Systems Research”. *Management Information Systems Quarterly* 28 (maaliskuu): 75–76. https://doi.org/10.2307/25148625.

Horiguchi, Yuji, Yukino Baba, Hisashi Kashima, Masahito Suzuki, Hiroki Kayahara ja Jun Maeno. 2017. “Predicting Fuel Consumption and Flight Delays for Low-Cost Airlines”. Teoksessa *AAAI.*

Johansson, Jesper M., Salvatore T. March ja David J. Naumann. 2003. “Modeling Network Latency and Parallel Processing in Distributed Database Design”. *Decision Sciences* 34 (4): 677–706. https://doi.org/https://doi.org/10.1111/j.1540-5414.2003.02409.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5414.2003.02409.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5414.2003.02409.x.

Kanevski, M. 2009. *Machine Learning for Spatial Environmental Data: Theory, Applications, and Software.* 1. painos. EPFL Press. https://doi.org/10.1201/9781439808085.

Karttunen, Hannu, Jarmo Koistinen, Elena Saltikoff ja Olli Manner. 2008. *Ilmakehä, sää ja ilmasto.* 353. ISBN: 978-952-5329-61-2.

Khanmohammadi, Sina, Salih Tutun ja Yunus Kucuk. 2016. “A New Multilevel Input Layer Artificial Neural Network for Predicting Flight Delays at JFK Airport”. Complex Adaptive Systems Los Angeles, CA November 2-4, 2016, *Procedia Computer Science* 95:237–244. ISSN: 1877-0509. https://doi.org/https://doi.org/10.1016/j.procs.2016.09.321.

Klein, Heinz K., ja Michael D. Myers. 1999. “A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems”. *MIS Quarterly* 23 (1): 67–93. ISSN: 02767783. http://www.jstor.org/stable/249410.

Korpela, Jari. 2019. “Koneoppimisen hyödyntäminen kaupallisen lentoyhtiön toiminnoissa”, http://urn.fi/URN:NBN:fi:jyu-201904242259.

Kärkkäinen, T., S. Äyrämö, T. Kilpeläinen ja K. Lahti, toimittaneet. 2003. “Data Mining: Osaraportit I, 2003.”

Laurence, Vorage. 2021. “Predicting Probabilistic Flight Delay for Individual Flights using Machine Learning Models”.

Lee, Hanbong, Waqar Malik, Bo Zhang, Balaji Nagarajan ja Yoon Jung. 2015a. “Taxi Time Prediction at Charlotte Airport Using Fast-Time Simulation and Machine Learning Techniques”. Kesäkuu. https://doi.org/10.2514/6.2015-2272.

———. 2015b. “Taxi Time Prediction at Charlotte Airport Using Fast-Time Simulation and Machine Learning Techniques”. Kesäkuu. https://doi.org/10.2514/6.2015-2272.

Lentosääoppia harrasteilmailijoile. 2020. 81–113. Traficom. Viitattu 10. toukokuuta 2022. https://www.traficom.fi/sites/default/files/media/file/Lentos%C3%A4%C3%A4oppia%20harrasteilmailijoille%202020-06-08_v2.pdf.

Lentosääpalvelut Suomessa. 2021. Ilmatieteen laitos. Viitattu 9. helmikuuta 2022. https://ilmailusaa.fi/pdf/Lentosaapalvelut_Suomessa_01-2021.pdf.

March, Salvatore T., ja Gerald F. Smith. 1995. "Design and natural science research on information technology". *Decision Support Systems* 15 (4): 251–266. ISSN: 0167-9236. https://doi.org/10.1016/0167-9236(94)00041-2. https://www.sciencedirect.com/science/article/pii/0167923694000412.

Mitchell, Rory, ja Eibe Frank. 2017. "Accelerating the XGBoost algorithm using GPU computing" (huhtikuu). https://doi.org/10.7287/peerj.preprints.2911.

Norin, Anna, Tobias Granberg, Di Yuan ja Peter Värbrand. 2011. "Airport logistics – A case study of the turn-around process". *Journal of Air Transport Management - J AIR TRANSP MANAG* 20 (tammikuu). https://doi.org/10.1016/j.jairtraman.2011.10.008.

Nosedal, Jenaro, ja Miquel Eroles. 2017. "Causal analysis of aircraft turnaround time for process reliability evaluation and disruptions' identification". *Transportmetrica B: Transport Dynamics* 6 (toukokuu): 1–14. https://doi.org/10.1080/21680566.2017.1325784.

Nykänen, Visa. 2017. "Päätöspuiden käyttö koneoppimisessa".

Official Journal of the European Union. 2008. THE COMMISSION OF THE EUROPEAN COMMUNITIES. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:254:0001:0238:En:PDF.

Oreschko, Bernd, Thomas Kunze, Michael Schultz, Hartmut Fricke, Vivek Kumar ja Lance Sherry. 2012. "Turnaround Prediction with Stochastic Process Times and Airport Specific Delay Pattern". Kesäkuu.

Oreschko, Bernd, Michael Schultz, Jenny Elflein ja Hartmut Fricke. 2010. "Significant Turnaround Process Variations due to Airport Characteristics" (kesäkuu): 1.

Pamplona, Daniel Alberto, Li Weigang, Alexandre Gomes deBarros, Elcio Hideiti Shiguemori ja Claudio Jorge Pinto Alves. 2018. “Supervised Neural Network with multilevel input layers for predicting of air traffic delays”. Teoksessa *2018 International Joint Conference on Neural Networks (IJCNN),* 1–6. https://doi.org/10.1109/IJCNN.2018.8489511.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel ym. 2011. “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research* 12:2825–2830.

Piatetsky-Shapiro, G. 1991. “Knowledge discovery in real databases: A report on the IJCAI-89 workshop”. *AI Magazine,* numero 11, 68–70.

Polikar, Robi. 2012. “Ensemble learning”. Teoksessa *Ensemble machine learning.* Springer.

Pyle, Dorian. 1999. *Data preparation for data mining.* 11. morgan kaufmann.

Ramírez-Gallego, Sergio, Bartosz Krawczyk, Salvador García, Michał Woźniak ja Francisco Herrera. 2017. “A survey on data preprocessing for data stream mining: Current status and future directions”. *Neurocomputing* 239:39–57. ISSN: 0925-2312. https://doi.org/https://doi.org/10.1016/j.neucom.2017.01.078. https://www.sciencedirect.com/science/article/pii/S0925231217302631.

Schmidt, Michael. 2017. “A review of aircraft turnaround operations and simulations”. *Progress in Aerospace Sciences* 92 (toukokuu). https://doi.org/10.1016/j.paerosci.2017.05.002.

SPSS Modeler: IBM SPSS Modeler V18.3.0 documentation. 2022, versio 18.3. IBM. Viitattu 14. helmikuuta 2022. https://www.ibm.com/docs/en/SS3RA7_18.3.0/pdf/spss-modeler-18.3.0-documentation.pdf.

“Suomen ilmailukäsikirja”. 2022. Viitattu 14. maaliskuuta 2022. https://www.ais.fi/ais/aip/fi/index.htm.

Thiagarajan, Balasubramanian, Lakshminarasimhan Srinivasan, Aditya Sharma, Dinesh Sreekanthan ja Vineeth Vijayaraghavan. 2017. “A machine learning approach for prediction of on-time performance of flights”, 1–6. Syyskuu. https://doi.org/10.1109/DASC.2017.8102138.

Tsichritzis, Dennis. 1997. “The dynamics of innovation”.

Wang, H., M. Wang ja Y. Wu. 2017. *Development of an Aircraft Turnaround Time Estimation Model Based on Discrete Time Simulation.* ITITS. https://doi.org/10.3233/978-1-61499-785-6-29.

Weibo, Liu, Wang Zidong, Liu Xiaohui, Zeng Nianyin, Liu Yurong ja E. Alsaadi Fuad. 2017. "A survey of deep neural network architectures and their applications". *Neurocomputing* 234:11–26. ISSN: 0925-2312. https://doi.org/https://doi.org/10.1016/j.neucom.2016.12.038. https://www.sciencedirect.com/science/article/pii/S0925231216315533.

Verleysen, Michel, ja Damien Francois. 2005. "The Curse of Dimensionality in Data Mining and Time Series Prediction". Teoksessa *Computational Intelligence and Bioinspired Systems,* toimittanut Joan Cabestany, Alberto Prieto ja Francisco Sandoval, 758–770. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-540-32106-4.

Wu, Cheng-Lung. 2010. "Airline operations and delay management: Insights from airline economics, networks and strategic schedule planning". *Airline Operations and Delay Management: Insights from Airline Economics, Networks and Strategic Schedule Planning* (tammikuu): 1–241.

Vuori, Petri, ja Katriina Ahlroth. 2022. "Puheenvuoro: Euroopassa lentoslotien käsittely on tarkkaan säänneltyä". *Uutishuone,* viitattu 18. tammikuuta 2022. http://www.eurocontrol.int/publications/european-airline-delay-cost-reference-values.

Yin, Jianan, Yuxin Hu, Yuanyuan Ma, Yan Xu, Ke Han ja Dan Chen. 2018. "Machine Learning Techniques for Taxi-out Time Prediction with a Macroscopic Network Topology". Syyskuu. https://doi.org/10.1109/DASC.2018.8569664.

Yu, Lei, ja Huan Liu. 2003. "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution", 2:856–863. Tammikuu.

Zhang, Dahai, Liyang Qian, Baijin Mao, Can Huang ja Yulin Si. 2018. "A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGBoost". *IEEE Access* PP (huhtikuu). https://doi.org/10.1109/ACCESS.2018.2818678.

Zhang, Yagang. 2010. *New Advances in Machine Learning.* 19. ISBN: 978-953-307-034-6.

Äyrämö, Sami. 2006. *Knowledge mining using robust clustering*. 63. University of Jyväskylä. ISBN: 951-39-2621-4. https://jyx.jyu.fi/handle/123456789/13286.

Liitteet

A Malli lumitiedon luokittelusta

Sääkoodi	Räntä	Lumi	Jäätävä
-SHSN	-	5	-
-IC	-	5	-
IC	-	5	-
SHSN	-	4	-
-SN	-	3	-
SN	-	2	-
DRSN	-	3	-
BLSN	-	2	-
+DRSN	-	2	-
+BLSN	-	1	-
+SN	-	1	-
SHSG	-	4	-
-SG	-	4	-
SG	-	3	-
-PL	-	4	-
SNPL	-	2	-
PL	-	3	-
-SHGS	-	4	-
SHGS	-	3	-
GS	-	2	-
GR	-	3	-

Taulukko 8. Lumen sääluokat 1

Sääkoodi	Räntä	Lumi	Jäätävä
-SHRASN	4	-	-
-SHSNRA	4	-	-
SHRASN	3	-	-
-DZSN	3	-	-
-SNDZ	3	-	-
-RASN	2	-	-
RASN	1	-	-
-SNRA	2	-	-
SNRA	1	-	-
+SNRA	1	-	-
-FZDZ	-	-	4
FZFG	-	-	3
FZDZ	-	-	3
-FZRA	-	-	2
-FZUP	-	-	2
-FZDZSN	-	-	3
-FZRASN	-	-	3
FZRA	-	-	1
FZUP	-	-	1

Taulukko 9. Lumen sääluokat 2

B Helsingin kiitotien ennustaminen

XGBClassifier-hyperparametrien arvot:

- subsample = 0.67,
- scale_pos_weight = 0.95,
- n_estimators = 776,
- min_samples_split = 9,
- min_samples_leaf = 1,
- min_child_weight = 0.42,
- max_features = "sqrt",
- max_depth = 32,
- learning_rate = 0.084,
- gamma = 0.17,
- colsample_bytree = 0.82,
- colsample_bylevel = 0.965