

MS-E2177 Operaatiotutkimuksen projektityöseminaari

Junien peruuntumistodennäköisyyksien hyödyntäminen veturinkuljettajien työvuoroluetteloiden suunnittelussa

Loppuraportti

25.5.2017

Tapio Hautamäki (projektipäällikkö)
Juhani Mutikainen
Johanna Piipponen

Sisällysluettelo

1	Johdanto	3
2	Tausta ja tavoitteet	3
3	Kirjallisuuskatsaus	4
3.1	Vastaava tutkimus.....	4
3.2	Logistinen regressiomalli	5
3.3	Aikasarjamallit.....	6
3.4	Päätöksenteko	6
4	Käytetyt tietoaineistot.....	7
5	Ennustemallit.....	8
5.1	Todennäköisyysmallinnusohjelman periaatteet.....	8
5.2	Keskiarvomalli	9
5.3	Viikonpäivämalli.....	10
5.4	Logistinen regressiomalli	10
5.5	Mallien hyvyyden arviointi.....	11
5.5.1	Ennusteen kokonaisvirheen arviointi.....	11
5.5.2	Ennusteiden visualisointi	12
6	Käyttöliittymä	16
6.1	Peruuntumistodennäköisyys työvuorolle	16
6.2	Peruuntumisten vaikutus työvuoroon	16
6.3	Todennäköisyyksien ja vaikutusten yhdistäminen	18
6.4	Päätöksentekotaulukko	18
6.5	Käyttöliittymän toiminta.....	21
6.5.1	Peruuntumistodennäköisyyksien laskenta	21
7	Yhteenveto	22
	Lähdeluettelo	23
	Liite: Itsearviointi	24

1 Johdanto

Tämä projektityö on tehty osana Aalto-yliopiston kurssia MS-E2177 Operaatiotutkimuksen projektityöseminaari. Projektityön tavoitteena on ollut tuottaa päätöksentekotukea veturinkuljettajien työvuorojen suunnittelijoille junien peruuntumistodennäköisyyksien avulla. Projektityön toteutuksesta on vastannut kolmen hengen opiskelijaryhmä ja se on toteutettu kevään 2017 aikana.

Projektityön aiheen asetti logistiikkakonserni VR Group, joka harjoittaa henkilö- ja tavaraliikennettä pääasiassa Suomen rautateillä ja maanteillä. VR:n junaliikennöinnin suunnitteluyksikkö on kehittänyt tehokkaita menetelmiä veturinkuljettajien työvuorojen automaattiseen järjestelyyn esimerkiksi tauotuksen ja lomien aiheuttamien rajoitteiden perusteella. Tavoitteena on ottaa huomioon yhä enemmän tekijöitä työvuoroluetteloiden koostamisessa. Junien peruuntumisen huomioiminen on rajoittunut henkilökunnan kykyyn muistaa ongelmalliset tapaukset ja järjestellä työvuoroja käsin peruuntumisherkkien junien huomioimiseksi. Projektityön aiheena on ollut tämän prosessin kehittäminen päätöksentekoa tukevan työkalun avulla.

Projektin lopputuloksena syntyi R-ohjelmointikielellä kirjoitettu ohjelma, joka tuottaa ennusteita junien peruuntumistodennäköisyydestä, ja järjestää työvuoroja tämän tiedon perusteella. Lisäksi ohjelma huomauttaa sellaisista työvuoron ominaisuuksista, jotka junien peruuntumisiin yhdistettynä joko lisäävät tai vähentävät peruuntumisen vaikutuksia. Ohjelman kehitystyön helpottamiseksi tuotimme myös työkaluja, joiden avulla uusien ennustemallien hyvyyden arviointi on mahdollista.

Seuraavassa luvussa käydään tarkemmin läpi työn tausta ja tavoitteet. Luvussa 3 tutustutaan työn aiheeseen liittyvään kirjallisuuteen ja luvussa 4 ohjelman käyttämiin tietoaineistoihin. Luvussa 5 esitellään todennäköisyysmallinnusohjelma ja sen käyttämät kolme ennustemallia. Lisäksi tässä luvussa esitellään ennustemallien hyvyyden arviointiin kehitetyt työkalut ja näiden avulla saadut arviot kehitetyistä ennustemalleista. Luvussa 6 esitellään ohjelman käyttöliittymä ja siihen rakennetut päätöksentekoa tukevat ominaisuudet. Raportti päättyy yhteenvetoon luvussa 7.

2 Tausta ja tavoitteet

Veturinkuljettajat VR:n junaliikennöintidivisioonassa ajavat sekä kaukoliikenteen henkilöjunia että tavarajunia. Yhteen työvuoroon voi kuulua sekä henkilö- että tavarajunia, ja yksi työvuoroluettelo on kolmen viikon pituinen. Työvuoroluettelot annetaan kuljettajille etukäteen ja luettelo on hankala tehdä muutoksia myöhemmin, vaikka muutoksille on usein tarvetta. Henkilöjunia perutaan harvoin, mutta tavarajunien kannalta tilanne on toinen: tavarajunia voidaan perua työvuoroluetteloiden laatimisen jälkeen tai niitä voi tulla lisää.

Osa veturinkuljettajista on "heittomiehiä", joille työvuoroluetteloita ei tarvitse julkaista viikkoja etukäteen vaan vasta edellisenä iltana. Nämä heittomiehet hoitavat ne ylimääräiset työvuorot, joille ei ole kuljettajia, ja heitä on enintään 17 % kuljettajista. Työvuoro jää heittomiehille ajettavaksi, kun

- a) vuoron oikea kuljettaja on yllättäen poissa,
- b) tulee tarve lisäjunien ajamiselle tai
- c) vuoroa ei ole suunniteltu työvuoroluettelo.

Näistä kaksi ensimmäistä tapausta tapahtuvat aina satunnaisesti. Sen sijaan päätös siitä, mitkä työvuorot kiinnitetään työvuoroluetteloihin ja mitkä jätetään heittoon, on suunnittelijoiden harkittavissa.

Työvuorojen sijoittaminen joko työvuoroluetteloihin tai heittoon vaikuttaa siihen, kuinka monta tuntia veturinkuljettajien työajasta käytetään hyödyksi. Veturinkuljettajien työaika kolmen viikon jaksolla on enintään 114 h 45 min. Syntyy tilanteita, joissa vuoron kuljettajien työtunnit jäävät täyttymättä junien peruuntumisten vuoksi, heittomiesten joustavuutta ei hyödynnetä riittävästi, ja lisäjunia joudutaan ajamaan pakon edessä ylitöinä. Tavoite on, että kiinteää vuoroa ajavat kuljettajat saavat maksimityötuntinsa täyteen eli junia ei peruuntuisi vuorosta. Toisaalta heittomiehet ajaisivat ne työvuorot, jotka ovat jääneet yli sekä lisäjunat, joista ei ollut tietoa työvuoroja laadittaessa.

Tämän projektin tavoitteena on ollut auttaa työvuorosuunnittelijoita päättämään, millaiset työvuorot tulisi sijoittaa työvuoroluetteloihin ja mitkä jätetään heittoon niiden sisältämien peruuntumisherkkien junien vuoksi. Tätä varten asiakkaalle on ollut tarkoituksena tuottaa analyttinen työkalu, joka jaottelee työvuorot kolmeen kategoriaan:

- a) 'tulee sijoittaa työvuoroluetteloihin',
- b) 'tulisi ehkä sijoittaa luetteloihin' ja
- c) 'ei ainakaan sijoiteta luetteloihin'.

Työkalulla on pyritty tuottamaan helppokäyttöistä tietoa erityisesti ehkä-kategorian vuoroista päätöksentekoa avustamaan. Tavoitteena on ollut myös analysoida, millaisia vaikutuksia yksittäisen junan peruuntumisella on työvuoron ajamiseen ja kannattaisiko vuoron niiden perusteella olla työvuoroluettelossa vai heitossa.

3 Kirjallisuuskatsaus

Tässä kappaleessa tutustutaan projektin teoreettiseen taustaan. Esittelemme projektimme aihealuetta vastaavaa aiempaa tutkimusta ja käsittelemme logistisen regressiomallin, aikasarjamallien sekä päätöksentekoanalyysin teoriaa.

3.1 Vastaava tutkimus

Projektimme ongelmakenttä on melko uniikki yhdistelmä ennustemallien luomista ja Suomen rautatieliikenteen erikoispiirteisiin liittyvää päätöksentekoanalyysiä. Emme onnistuneet löytämään kirjallisuuskatsauksessamme kovin hyvin vastaavia case-esimerkkejä. Esittelemme kuitenkin tässä kappaleessa joitakin otteita kirjallisuudesta liittyen joko peruuntumisten ennustamiseen tai työvuoroluetteloiden luomiseen.

Asiakkaiden tekemien peruuntumisten mallinnuksesta on tuoretta tutkimusta. Neuroverkkojen avulla on kyetty luomaan toimivia ennusteita tässä kontekstissa (Huang et al. 2013). Terveystieteiden tutkimuksessa puolestaan saatiin hyviä tuloksia mallintamalla peruuntumisia logistisen regressiomallin ja bayesiläisen päättelyn yhdistelmällä (Alaeddini et al. 2015). Tässä tutkimuksessa otettiin myös ansiokkaasti huomioon eri potilastapaamisia painottamalla peruuntumisen haittavaikutus, joka riippuu mm. viikonpäivästä ja klinikkatyypistä. Hieman vanhemmassa väitöstutkimuksessa käytettiin elinaika-analyysiä lentomatikustajien perumiskäyttötymisen mallinnukseen (Iliescu 2008).

Työvuoroluetteloiden optimointia on tutkittu monilla eri aloilla. Hollannissa on käytetty operaatiotutkimuksen menetelmiä veturinkuljettajien työvuoroluetteloiden luomiseen, mutta junien peruuntumisen huomioimista ei ole mainittu (Abbink et al. 2005; Alfieri et al. 2007). Veturinkuljettajien työvuoroluetteloiden optimoinnissa on keskitytty työsopimusten ja esimerkiksi lomien aiheuttamien rajoitteiden huomioimiseen sekä tyhjän ajan minimointiin (ks. esim. Hanne et al. 2009; Lezaun et al. 2007). Emme kuitenkaan löytäneet mainintoja junien peruuntumisten huomioinnista tässä yhteydessä. Näiden rautatieliikenteen esimerkkien lisäksi vanhempaa tutkimusta työvuoroluetteloiden luomisesta yleisesti on koottu myös kirjallisuuskatsaukseksi (Van den Bergh et al. 2013). Tietullien tarkastajien työvuoroluetteloihin ja reititykseen keskittyneessä tuoreessa saksalaistutkimuksessa käytettiin kokonaislukuoptimointia luetteloiden luomiseen ja lisäksi kaksitavoitearviointia tasapainoiluun luettelon hyväksynnän ja valvonnan tehokkuuden välillä (Borndörfer et al. 2016). Lineaarista sekalukuoptimointia on puolestaan käytetty hiljattain onnistuneesti lääkäreiden työvuorojen suunnitteluun (Fügener et al. 2015).

3.2 Logistinen regressiomalli

Logistisella regressiomallilla mallinnetaan todennäköisyyttä sille, että kategorinen selitettävä muuttuja Y kuuluu tiettyyn luokkaan. Oletetaan luokkia olevan kaksi: 0 ja 1. Selitettävä muuttuja Y noudattaa siis Bernoulli-jakaumaa. Oletetaan, että todennäköisyyteen $P(X)$ vaikuttaa p selittävää muuttujaa eli X_1, \dots, X_p . Logistinen regressiomalli on muotoa

$$P(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$$

jossa β_0, \dots, β_p ovat mallin parametreja. Mallin parametrit estimoidaan usein suurimman uskottavuuden menetelmällä. Selittävät muuttujat voivat olla jatkuvia tai kategorisia, jolloin malliin voidaan syöttää dummy-muuttujia avuksi. Selittävien muuttujien tulee olla toisistaan riippumattomia kuten myös havaintojen.

Logistinen regressiomalli on osa yleistettyjen lineaaristen mallien perhettä, sillä yllä oleva yhtälö voidaan esittää myös muodossa

$$\log\left(\frac{P(X)}{1 - P(X)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p.$$

Yhtälön vasenta puolta voidaan kutsua mallin logitiksi.

Logistisen regression ennustama todennäköisyys tulkitaan niin, että mikäli $P(X) > 0,50$, havainto kuuluu luokkaan 1 ja muulloin luokkaan 0. Logistinen regressiomalli tuottaa aina S-käyrän, jonka arvojoukko on välillä $[0, 1]$, joten tulokset ovat mielekkäitä todennäköisyyden mielessä. (James et al. 2013)

Logistinen regressiomalli on yksi monista luokittelu- eli klassifointimenetelmistä. Muita menetelmiä ovat muun muassa klassifointipuut (classification trees) ja tukivektorikone (support vector machines).

3.3 Aikasarjamallit

Yksi tapa ennustamiseen on käyttää aikasarjamalleja. Yleisiä ovat niin sanotut ARMA-mallit (autoregressive-moving-average) eli autoregressiiviset liukuvan keskiarvon mallit (Choi 2012). ARMA(p, q)-mallissa muuttujan X arvo indeksillä t voidaan esittää muodossa

$$X_t = c + \epsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i},$$

missä c on vakio, ϵ on toteuman poikkeama mallin ennusteesta eli ns. virhetermi ja φ sekä θ painokertoimina toimivia vakioita. Malli siis käytännössä ennustaa menneiden toteumien ja virhetermien painotettujen keskiarvojen avulla. Arvoilla p ja q voidaan valita, kuinka montaa mennyttä toteumaa käytetään ennustamiseen. Malli sovitetaan saatavilla olevaan dataan, minkä jälkeen sillä voi ennustaa tulevia arvoja. (Choi 2012)

ARMA-malleja ei voida suoraan käyttää tähän ongelmaan, sillä ne on tarkoitettu jatkuville muuttujille. Ongelmassa ennustukseen käytettävä data on binääristä: 0, jos junavuoro toteutuu ja 1, jos se perutaan. On kuitenkin olemassa myös monimutkaisempia ja yleiskäyttöisempiä malleja, jotka toimivat myös diskreettien muuttujien kanssa. R-ohjelmointikielelle on saatavilla ns. GLARMA-malleja (generalized linear autoregressive moving average) toteuttava glarma-paketti. (CRAN 2017) Tämänkin toteutuksessa on kuitenkin ongelmana se, että projektissa ei varsinaisesti haluta ennustaa, peruuntuuko juna tietynä päivänä, vaan sitä, millä todennäköisyydellä juna peruuntuu. Muuten junien ja niistä koostuvien työvuorojen vertailu on vaikeaa. Varsinaisia aikasarjamalleja ei projektissa lopulta toteutettu.

3.4 Päätöksenteko

Projektin keskeinen tutkimusongelma voidaan muotoilla vastauksen etsimisenä kysymykseen: Kuinka suuri on työvuoron sisältämien työtehtävien toteutuva kesto niiden suunniteltuun keston verrattuna? Joidenkin ajotehtävien peruuntuessa työtä ei pystytä juurikaan uudelleenjärjestelemään, jolloin työn kesto ei juuri muutu. Toisinaan taas junan peruuntuminen vapauttaa kuljettajan muihin tehtäviin. Odotusarvoiseen keston vaikuttaa työvuoron junien peruuntumistodennäköisyydet ja muitakin tekijöitä. Yleinen lähestymistapa tällaisessa tilanteessa on muodostaa vaikuttavista tekijöistä arvofunktio, joka kuvastaa päätöksentekijän tavoitteita. Mitä suuremman arvon arvofunktio saa, sitä mieluisampi tilanne on päätöksentekijälle. Yleisin arvofunktio on additiivinen arvofunktio, joka voidaan esittää muodossa

$$V_i = \sum_j w_j x_{i,j},$$

missä V_i on vaihtoehdon i arvo päätöksentekijälle, w_j on yhden vaikuttavan tekijän painoarvo ja $x_{i,j}$ on yhden tekijän arvo vaihtoehdolla i . (Belton 1986)

Jos vaikuttavat tekijät ovat tiedossa, haasteena on selvittää tilanteeseen sopivat painokertoimet (w_j), jotka kuvastaisivat päätöksentekijän preferenssejä (tässä tapauksessa vaikutusta hyödyllisen työn keston). Tämä pitää suorittaa päätöksentekijää haastatteleamalla. Tapoja tähän on useita. Päätöksentekijää voi esimerkiksi pyytää valitsemaan kuvitteellisia eriäviä vaihtoehtoja, jotka olisivat keskenään yhtä mieluisia. (Torrance et al. 1982)

4 Käytetyt tietoaineistot

Projektissa on käytetty pääosin kahta aineistoa: työvuoroaineistoa ja menneiden junien aikataulu- ja toteumatietoja. Työvuoroaineisto on saatu VR:ltä ja se kattaa ma 23.1. - su 13.2.2017 välisen ajanjakson. Aineistossa tieto kaikista tämän ajanjakson työvuoroista on koottu yhteen .csv-taulukkoon. Taulukon rivit ovat yksittäisiä työtehtäviä ja sarakkeet sisältävät yksityiskohtaista tietoa kustakin työtehtävästä. Sarakkeet kertovat esimerkiksi, mihin työvuoroon ja mille ajanjaksolle työtehtävä kuuluu.

Aikataulu- ja toteumatiedot ovat Liikenneviraston jakamaa avointa dataa, ja sitä on saatavilla 8.3.2015 alkaen (Liikennevirasto 2017). Tietoja voi ladata Liikenneviraston kehittämästä Digitraffic-palvelusta avoimen rajapinnan kautta. Tietoja säilytetään noin 720 päivän ajan ja aikataulutietoja on lisäksi tarjolla 10 päivää tulevaisuuteen. Rajapinta on selkeästi dokumentoitu ja apuna on myös aktiivinen keskusteluryhmä. Palvelua kehitetään jatkuvasti. (Liikennevirasto 2017)

Tämän projektin kannalta tärkeintä on hakea kaikkien junien historiatiedot päivämääräkohtaisesti. Kaikki rajapinnan jakamat tiedot tulevat JSON-tiedostomuodossa, jota R osaa tulkita jsonlite-paketin avulla. Hakuun voidaan ottaa mukaan vain aikataulutiedot tai niiden lisäksi myös toteumatiedot eli tieto siitä, mihin aikaan juna todellisuudessa saapui asemille. Tässä projektissa haetaan sekä aikataulu- että toteumatiedot. Toteumatietoja ei hyödynnetä, mutta jatkokehitystä varten myös junien myöhästymisen voidaan näiden tietojen avulla ottaa huomioon.

Seuraavassa on esimerkki datan hausta ja sisällöstä:

```
library(jsonlite)
yhden.paivan.junat <- fromJSON("https://rata.digitraffic.fi/
  api/v1/history?departure_date=2017-01-23")
```

Haku palauttaa data.frame-objektin, jolla on niin monta ulottuvuutta kuin junia kyseisenä päivänä. Päivämäärän perusteella haetaan ne junat, joiden lähtöpäivä ensimmäiseltä asemalta osuu kyseiselle päivälle. Aikataulutietojen kellonajat ovat UTC-aikavyöhykkeen mukaiset.

Tämän työn kannalta tärkeimmät kentät ovat trainNumber ja cancelled. Edellinen kertoo junanumeron väliltä 1—99999. Jälkimmäinen kenttä on TRUE, jos junan peruminen on tehty 10 vuorokauden sisällä. Yli 10 vuorokautta sitten peruttuja junia ei palauteta rajapinnasta lainkaan. Muussa tapauksessa kentän arvo on FALSE eli junaa ei peruttu. (Liikennevirasto 2017) Junakohtainen peruminen on määritelty siten, että juna ei ole peruttu, mikäli se pääsi matkaan lähtöasemaltaan. Kyseinen cancelled-kenttä palautetaan myös jokaiselta väliasemalta erikseen, jos halutaan tarkastella osaväliperumisia. Tämän työn rajauksiin kuitenkin kuului, ettei osaväliperuuntumisia oteta huomioon.

Päivämäärän, junanumeroiden ja perumistiedon perusteella alustettiin taulukko, jonka sarakkeina ovat kaikki mahdolliset tavarajunanumerot väleiltä [2000, 5999] ja [20000, 59999] eli yhteensä 44000 saraketta. Rivit kuvaavat havaintoja ajalta 8.3.2015—3.3.2017 eli perumishavaintoja on 727 päivältä. Taulukon arvoiksi kirjoitettiin

- 1, kun cancelled-kentän arvo oli TRUE;
- 0, kun cancelled-kentän arvo oli FALSE; ja
- NA (alustus), kun junaa ei pystytty palauttamaan rajapinnasta.

Hieman yksinkertaistettuna yllä olevat luokitukset voidaan määrittellä myös junan peruuntumisena, junan ajamisena ja sellaisena päivänä, ettei junaa oltu edes suunniteltu ajettavaksi.

Rajapinnan käyttöä on rajoitettu viiteen pyyntöön sekunnissa (Liikennevirasto 2017) ja suurien tietomäärien haku on siksi hidasta. Sen vuoksi historiatiedot ladattiin kiintolevyille. Tiedot väliltä 8.3.2015—3.3.2017 vievät noin 11 Gt levytilaa. Historiatiedoista muodostettu yksinkertaistettu taulukko on vain noin 540 kt R-ohjelmointikielen lukemassa rds-muodossa ja noin 94 Mt useimpien taulukko-ohjelmien lukemassa csv-muodossa.

5 Ennustemallit

Projektissa toteutettiin R-ohjelma mallintamaan junien peruuntumistodennäköisyyttä ja antamaan tukea päätöksenteossa. Tässä kappaleessa esitellään todennäköisyyden mallinnukseen luotu ohjelma ja sen käyttämät mallit. Mallien teoreettinen tausta on esitelty aiemmissa kappaleissa, mutta tässä esitämme, miten niitä hyödynnettiin käytännössä. Lisäksi vertaillaan mallien hyvyttä. Malleja on kolme: keskiarvomalli, viikonpäivämalli ja logistinen regressiomalli.

5.1 Todennäköisyydsmallinnusohjelman periaatteet

Todennäköisyyden mallintaminen on suunniteltu niin, että sitä olisi mahdollisimman sujuvaa käyttää työvuoroaineiston kanssa, mutta myös mallien kattava testaaminen onnistuisi. Siksi todennäköisyyden mallintaminen on jaettu kahteen funktioon. CalculateProbabilities-funktio on alemman tason funktio, joka ennustaa annetuille junanumeroille peruuntumistodennäköisyyden useampana päivänä kerrallaan. Tuloksena on siis peruuntumistodennäköisyystaulukko, jossa sarakkeina on ennustettavat junanumerot ja riveinä päivämäärät, joille ennustus halutaan.

PredictProbabilitiesForEsalli-funktio käyttää edellistä funktiota, mutta vain yksittäisille päivämäärille kerrallaan. Funktiolle annetaan kaksi samanpituista vektoria, joista toisessa on junanumeroita ja toisessa päivämääriä, ja funktio ennustaa peruuntumistodennäköisyyden aina yhdelle junanumerolle yhtenä päivänä. Tuloksena on vektori, jonka alkiot ovat peruuntumistodennäköisyyksiä samalla indeksillä sijainneelle junanumerolle ja päivämäärälle.

Työvuoroaineisto sisältää työtehtäväsarakeen. Ajettavien junien lisäksi sarake sisältää myös muita työtehtäviä. Projektissa koodattiin siksi funktio, joka tulkitsee työtehtävien joukosta junanumerot ja alustaa muut tehtävät NA:ksi, jotta niille ei yritettäisi virheellisesti laskea peruuntumistodennäköisyyttä. VR:n kanssa käytyjen keskusteluissa määritettiin, että työtehtävä on junanumero, mikäli se alkaa numerosarjalla. Tämän säännön perusteella ohjelma tulkitsee työtehtäviä taulukon 1 mukaisesti.

Taulukko 1: Esimerkki työvuoroaineiston tehtäväsarakeen siistimisestä

Syöte	Siistitty
2030	2030
2030X	2030
02030	2030
X2030	NA
X2030Y	NA
ABC	NA
HKSKE.M.2	NA
CONNECT.9685-9686	NA

Tässä projektissa ollaan kiinnostuneita nimenomaan tavarajunien perumisista. VR vahvisti, että tavarajunat ovat sellaisia junia, joiden numero on välillä [2000, 5999] tai [20000, 59999]. Lisäksi selvitettiin, että välillä [52000, 55999] olevat tavarajunat ovat todellisuudessa samoja junia kuin välillä [2000, 5999] olevat, mutta ne on esimerkiksi väliaikaisesti peruttu ja laitettu uudelleen ajoon. Tämän merkiksi junanumeron eteen on laitettu luku 5 eli esimerkiksi juna 2030 on muuttunut junaksi 52030. Tämän vuoksi ennustuksessa käytetty data jatkokäsiteltiin niin, että välillä [2000, 5999] olevien junien peruuntumistieto (1, 0, NA) korjattiin ei-perutuksi (0), mikäli välillä [52000, 55999] ollut juna oli ei-peruttu (0). Sen jälkeen junia [2000, 5999] koskevat sarakkeet kopioitiin junille [52000, 55999] ristiriitojen välttämiseksi.

Junan peruuntumistodennäköisyyttä määrittäessä juna kuuluu yhteen neljästä luokasta junanumeron perusteella:

1. Mikäli junanumero on NA eli syötettä ei ole tulkittu oikeaksi junatehtäväksi, peruuntumistodennäköisyys on NA
2. Mikäli junanumero ei vastaa tavarajunaa eli ei ole välillä [2000, 5999] tai [20000, 59999], peruuntumistodennäköisyys on 0 eli juna ajetaan aina. Tähän luokkaan kuuluvat henkilöjunat, veturiajot ja ratatyökoneet.
3. Mikäli juna on tavarajuna, mutta sille ei ole saatavilla historiadataa, sen peruuntumistodennäköisyys on NA.
4. Mikäli juna on tavarajuna ja sille löytyy oma sarake, sille lasketaan peruuntumistodennäköisyys käyttäen yhtä kolmesta pian esiteltävästä todennäköisyysmallista. Todennäköisyysmalli laskee junalle peruuntumistodennäköisyyden väliltä [0,1]. Todennäköisyyden laskemisessa voi olla satunnaisesti ongelmia, jolloin todennäköisyydeksi voidaan myös palauttaa NaN (erityisesti nollalla jaettaessa) tai NA.

Esitellään seuraavaksi todennäköisyysmallit, joita mallinnusohjelma käyttää. Vaikka tässä työssä mallien lukumäärä on rajattu kolmeen, voi mallinnusohjelmaan lisätä malleja jatkokehitystarpeista riippuen, kunhan rajapinta mallinnusohjelman ja mallin välillä säilyy samana.

5.2 Keskiarvomalli

Yksinkertaisin tapa ennustaa tavarajunan peruuntumistodennäköisyys on laskea, kuinka suuri osuus kyseisen junanumeron aiemmista ajoista on peruttu. Saadaan välillä [0, 1] oleva luku, jolla voidaan suoraan estimoida todennäköisyyttä. Matemaattisesti tämä yleinen keskiarvomalli voidaan esittää siis muodossa

$$p_i = \frac{\sum i_{peruttu}}{\sum i_{suunniteltu ajoon}},$$

missä i on kyseinen junanumero. Malli voi käyttää todennäköisyyden laskemiseen joko kaikkea saatavilla olevaa dataa tai dataa vain tietyltä aikaväliltä, esimerkiksi viimeiseltä kolmelta kuukaudelta, jos halutaan käyttää vain viimeaikaista tietoa.

Yleisessä keskiarvomallissa ei siis huomioida, mille viikonpäivälle tai kuukaudelle ollaan ennustamassa, vaan oletetaan, että tietyllä junanumerolla peruuntumistodennäköisyys on lähes vakio ennustushetkestä lähtien suunniteltavan työvuoroluettelon päättymispäivämäärään asti eli päivittäin yli neljän viikon ajan. Tämän oletuksen yksinkertaistavuus on mallin suurin puute. Mahdollista trendiä keskiarvoisessa

peruuntumistodennäköisyydessä voi yrittää huomioida käyttämällä vain viimeaikaista dataa ennustamisessa, mutta tällöin toteumien vähyys heikentää keskiarvojen tarkkuutta.

Yleisen keskiarvomallin etuna on sen helppo ja nopea laskenta. Junanumerokohtaisen mallin sovittamiseksi riittää kahden summan jakolasku. Laskennassa ei myöskään tarvitse erotella junanumeroita ennustettavan päivämäärän mukaan, mikä nopeuttaa käyttöliittymää huomattavasti. Tätä mallia päädyttiin käyttämään toistaiseksi työkalun oletusasetuksena. Tämä seuraa todennäköisyysmallien yhdistämisestä työvuoroaineiston kanssa. Tätä käsitellään tarkemmin kappaleessa 6.5.1.

5.3 Viikonpäivämalli

Viikonpäiväkeskiarvomalli on kehittyneempi versio yleisestä keskiarvomallista. Malli toimii muuten täysin samoin kuin yleinen keskiarvo, mutta pelkän junanumeron sijaan lasketaan todennäköisyys jokaiselle viikonpäivälle erikseen. Ennustaessa juna numero x maanantaille lasketaan todennäköisyys siis pelkästään junan x maanantaisista peruuntumisista. Matemaattisesti viikonpäiväkeskiarvomalli voidaan esittää siis muodossa

$$p_{i,t} = \frac{\sum i_{\text{peruttu päivänä } t}}{\sum i_{\text{suunniteltu ajoon päivälle } t}},$$

missä i on kyseinen junanumero ja t viikonpäivä.

Viikonpäiväkeskiarvon etu yleiseen keskiarvomalliin nähden on se, että mahdollinen viikonpäiväkohtainen vaihtelu peruuntumisissa voidaan tunnistaa. Mikäli vaihtelu ei kuitenkaan ole merkittävää, mallista ei ole lisähyötyä. Käyttöliittymän kannalta viikonpäiväkeskiarvomalli on hitaampi kuin yleinen keskiarvomalli, sillä junanumerot pitää erotella päivämäärän mukaan. Käytännön ongelmana on lisäksi se, että työvuoroluettelossa ei ole varsinaisesti junien lähtöpäivämääriä vaan työvuorojen aloituspäivämääriä. Yöjunien takia nämä eivät ole läheskään aina samoja. Oikean viikonpäivän selvittäminen kaikissa tapauksissa on siis yllättävän hankalaa.

5.4 Logistinen regressiomalli

Logistista regressiomallia varten luotiin kaksi selittävää muuttujaa: kuukausi ja viikonpäivä. Selittävät muuttujat olivat kategorisia ja määrittävät, mikä kuukausi (tammikuu, helmikuu, ..., joulukuu) ja mikä viikonpäivä (maanantai, tiistai, ..., sunnuntai) oli kyseessä. Selittävät muuttujat oli helppo muodostaa päivämäärien avulla. Myös muita selittäviä muuttujia voidaan luoda ja käyttää. Tällaisia muuttujia ovat esimerkiksi pyhäpäivä (0/1), lähteekö juna kylmäasemalta (0/1), saapuuko juna kylmäasemalle (0/1) tai lumen syvyys ([0, Inf]). Tärkeintä on, että selittävän muuttujan ajatellaan vaikuttavan junan perumiseen. Vaikutus voidaan toki myös varmistaa testausvaiheessa.

Malli estimoitii jokaiselle junanumerolle erikseen kuten muidenkin mallien tapauksessa. Logistisen regressiomallin soveltaminen on rajatumpaa kuin keskiarvo- ja viikonpäivämallien, sillä logistinen regressiomalli on hyvin tarkka siitä, että jokaiselle ennustettavalle pisteelle tulee löytyä vertailupiste, jolla on samat ominaisuudet selittämissä muuttujissa. Eli mikäli tietty junanumero on koko historiansa ajan kulkenut maanantaista torstaihin, mutta yhtäkkiä muuttuikin kulkemaan maanantaista perjantaihin, malli ei löydä uudelle havainnolle vertailukohtaa, sillä "perjantai" ei ole aiemmin ollut osa selittävän muuttujan "viikonpäivä" arvojoukkoa.

Logistisen regressiomallin etu keskiarvo- ja viikonpäivämalliin on, että se ottaa huomioon keskiarvoisen perumisasteen ja viikonpäivän lisäksi myös kuukauden. Näin saadaan esiin esimerkiksi se, jos juna

perutaan tavallista enemmän esimerkiksi kesä- tai talvikuukausina kevääseen ja syksyyn verrattuna. Toisaalta malli on epärobustimpi kuin muut juuri siinä mielessä, että se tarvitsee paljon havaintoja jokaisesta viikonpäivä-kuukausi-yhdistelmästä muodostaakseen vakuuttavan ennusteen.

5.5 Mallien hyvyyden arviointi

Ennustemallien hyvyyden arvioimiseksi toteutimme joukon funktioita, jotka joko visualisoivat ennusteita tai laskevat tunnuslukuja. Tässä kappaleessa käymme ne läpi ja esittelemme niiden tuottamat tulokset kolmelle yllä olevalle mallille. Mallien hyvyyden arviointi toteutettiin jakamalla peruuntumisdata kolmeen jaksoon. Ensimmäistä osaa käytettiin ennustemallien syötteenä. Toinen jakso jätettiin tarkastelusta pois ja ennuste luotiin viimeiselle jaksolle. Tämä vastaa todellista tilannetta, jossa ennuste luodaan kolmen viikon suunnittelujaksolle ennen jakson alkua. Vertaamalla luotua ennustetta viimeisen jakson toteumaan pystyttiin saamaan käsitys eri mallien suorituskyvystä. Ennusteet luotiin kaikille projektiryhmän käytössä olleen työvuoroluettelon sisältämille junille. Näitä oli yhteensä 485 ja tarkasteltavalle kolmen viikon ajanjaksolle osui yhteensä 5100 junavuoroa.

5.5.1 Ennusteen kokonaisvirheen arviointi

Ennustamisessa tehdyn kokonaisvirheen arviointiin toteutettiin kolme menetelmää: mediaani-, keski- ja luokitteluvirhe. Ensimmäiset kaksi menetelmää toimivat samalla periaatteella. Ennustevirheet lasketaan erotuksena ennustejakson toteumasta ja ennustetuista peruuntumistodennäköisyyksistä. Toteuma on joko 0 (juna ajettiin) tai 1 (juna peruttiin). Erotuksena saadut ennustevirheet ovat siis välillä $[-1,1]$. Ennustevirheistä laskettiin joko mediaani tai keskiarvo. Näille tunnusluville laskettiin luottamusvälit bootstrap-menetelmällä (James et al. 2013).

Luokitteluvirheen laskemiseksi käyttäjän tulee määrittää luokkaraja, johon peruuntumistodennäköisyyttä verrataan. Jos peruuntumistodennäköisyys on alle tämän rajan, juna oletetaan ajettavan, ja muussa tapauksessa sen oletetaan peruuntuvan. Luokitteluvirhe lasketaan tällä menetelmällä väärin luokiteltujen junavuorojen osuutena kaikista junavuoroista. Käytimme omissa tarkasteluissamme 50 prosentin luokkarajaa.

Taulukkoon 2 on koottu tunnusluvut ja luottamusvälit eri mallien tuottamille ennusteille. Vertailun vuoksi taulukkoon on lisätty myös rivi ”ei peruuntumisia”. Tällä rivillä on käytetty ennustetta, jonka mukaan junat eivät peruunnu koskaan eli peruuntumistodennäköisyys on asetettu nolnaan kaikille päville.

Taulukon sarake ”Junavuorojen lkm” kertoo, kuinka monelle junavuorolle ennusteen ja toteuman vertailu tehtiin onnistuneesti. Tämä vertailu epäonnistuu yksittäisen junavuoron kohdalla, jos ennustetta tai toteumaa ei ole saatavilla kyseiselle päivälle. Viimeisellä rivillä on mukana kaikki junavuorot, koska siinä peruuntumistodennäköisyys asetettiin manuaalisesti nolnaan kaikkien päivien kohdalla. Näin ennuste oli saatavilla aina kun toteumakin oli saatavilla. Keskiarvomallin kohdalla ennustetta ei ollut saatavilla aivan jokaiselle junavuorolle. Tämä johtuu siitä, että jotkin junat ovat kulkeneet alle kuusi viikkoa ja näin ollen aineiston alkuosasta ei pystytty laskemaan keskiarvoa. Viikonpäivämallin kohdalla vertailu epäonnistuu hieman useammin. Nyt epäonnistumisia aiheuttavat myös junat, jotka on laitettu hiljattain kulkemaan sellaisena viikonpäivänä, jona ne eivät ole aiemmin kulkeneet. Logistinen regressiomalli pystyy tuottamaan ennusteen kaikkein harvimmin. Se epäonnistuu aina, kun juna ei ole historiansa aikana ennen kulkenut samassa kuussa ja samalla viikonpäivällä.

Taulukko 2: Tunnusluvut ja luottamusvälit eri mallien tuottamille ennusteille

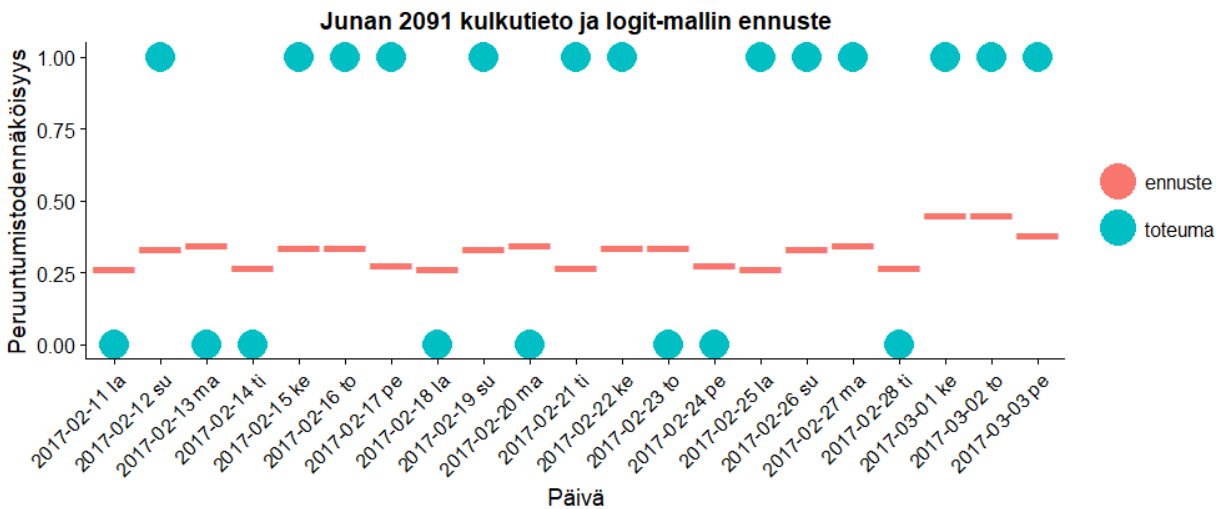
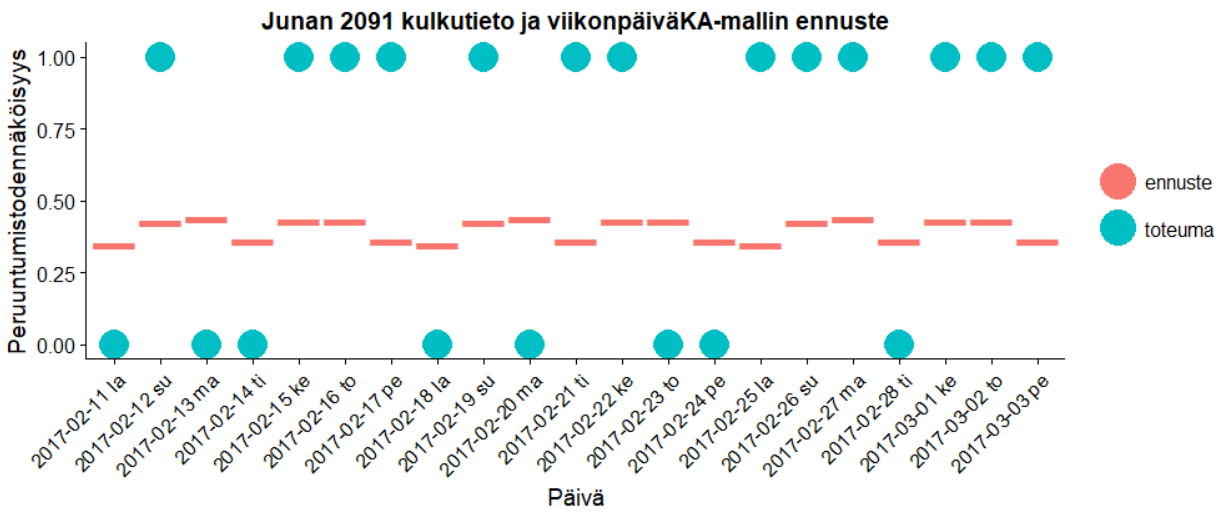
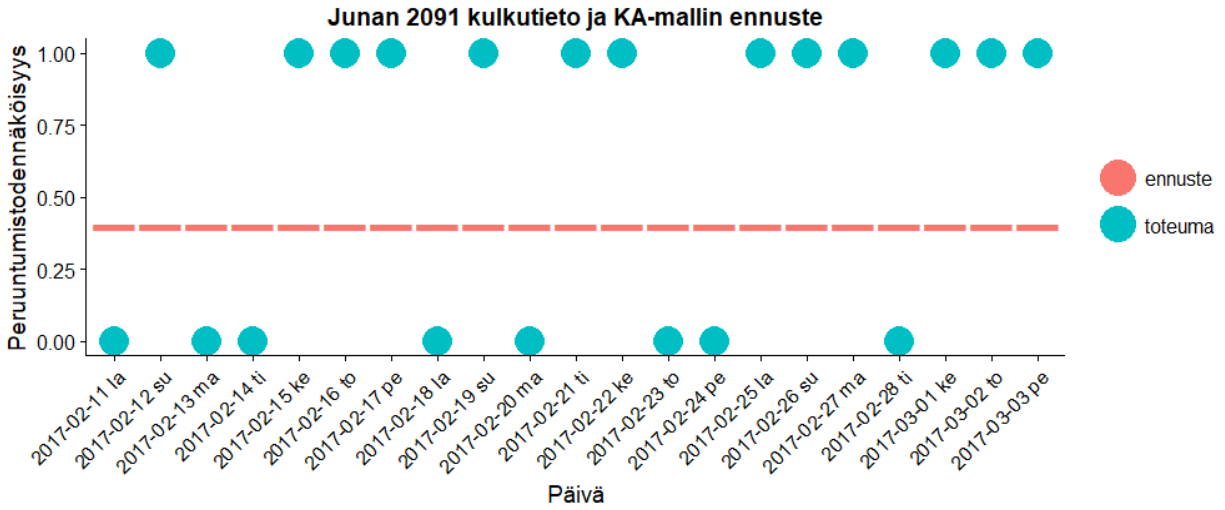
Ennustemalli	Junavuorojen lkm	Mediaanivirhe (95% luottamusväli)	Keskivirhe (95% luottamusväli)	Luokitteluvirhe, 50%:n luokkaraja
Keskiarvo	5068	-0,094 (-0,095; -0,093)	-0,027 (-0,036; -0,019)	14,52%
Viikonpäivä	5001	-0,075 (-0,077; -0,068)	-0,027 (-0,034; -0,021)	14,86%
Logistinen regressio	4348	-0,022 (-0,025; -0,018)	0.006 (-0,003; 0,015)	14,49%
Ei peruuntumisia	5100	0 (0; 0)	0.142 (0,132; 0,151)	14,16%

Mediaani- ja keskivirheen perusteella logistinen regressiomalli on selkeästi paras, vaikka kaksi huomattavasti yksinkertaisempaa mallia pärjäävät hyvin. Luokitteluvirhe ei vaikuta olevan järkevä ennusteen hyvyyden mittari ainakaan 50%:n luokkarajalla, sillä olettamalla kaikki peruuntumistodennäköisyydet nolaksi päästään lähes samaan lukemaan kuin kaikilla kolmella ennustemallilla. Luokitteluvirheen tarkastelu ei myöskään ollut projektissa kovin olennaista, koska käytännössä eri junavuoroille laskettujen todennäköisyyksien keskinäinen järjestys vaikuttaa päätöksentekoon, eikä tietyn rajan ylittäminen.

5.5.2 Ennusteiden visualisointi

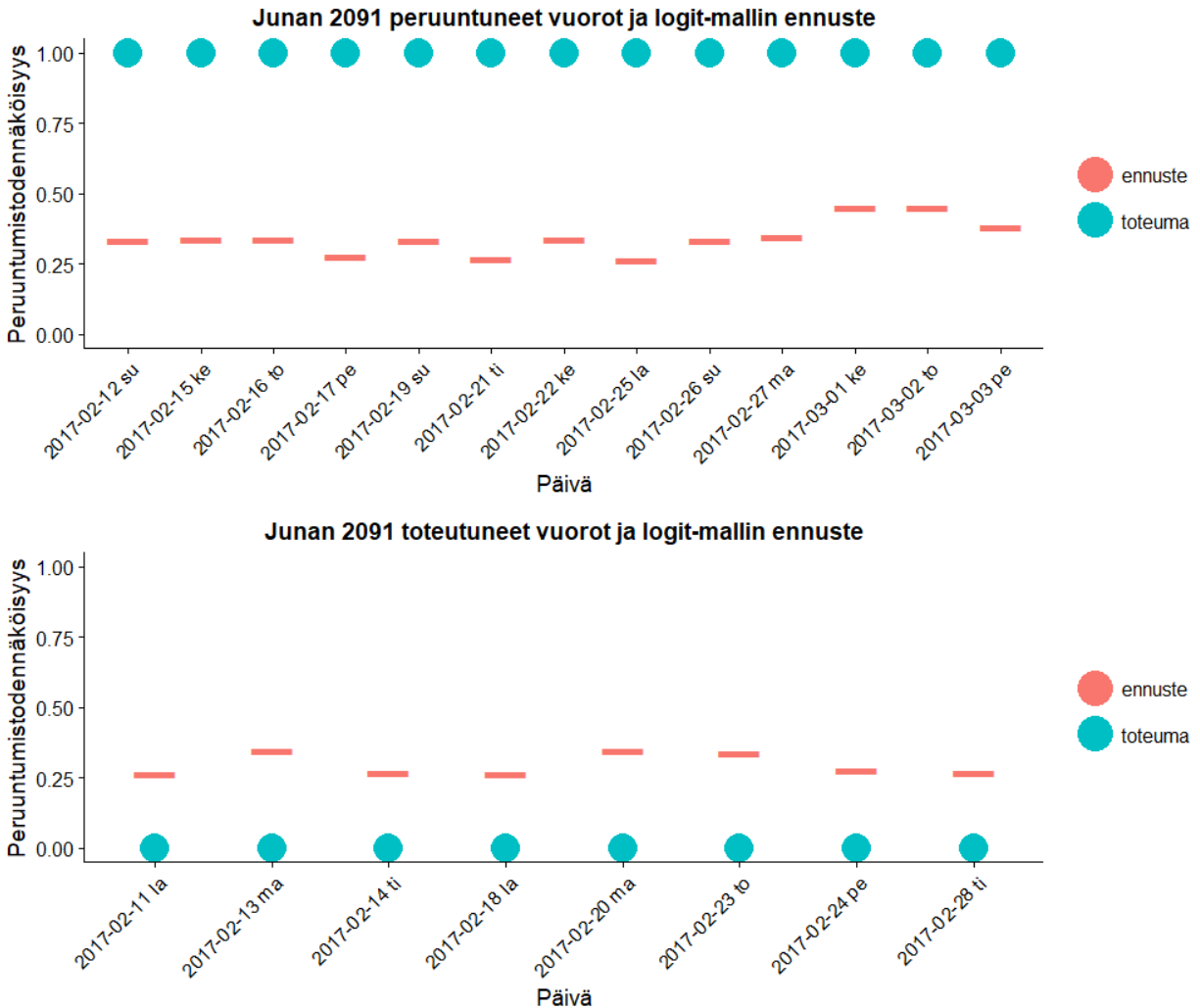
Ennusteiden visualisoimiseksi luotiin kaksi menetelmää. Ensimmäisellä menetelmällä voidaan luoda kuvaajia yhden junan ennusteista ja toisella menetelmällä visualisoidaan kaikkia ennustettuja junavuoroja kerralla. Yhdelle junalle voidaan piirtää aikasarjakuvaaja, jossa näkyy jokaiselle päivälle ennustettu peruuntumistodennäköisyys sekä kulkutieto (peruuntunut tai ajossa). Kuvassa 1 näkyy nämä kuvaajat piirrettyinä junalle 2091 ja kaikille kolmelle ennustemallille. Juna 2091 on Sköldvikistä Vainikkalaan kulkeva aamuöinen tavarajuna, joka perutaan usein. Toteuman arvo 1 kuvaa sitä, että juna peruuntui, ja toteuman arvo 0 sitä, että juna ei peruuntunut.

Kuvasta 1 nähdään, miten monimutkaisempien mallien tuottamat ennusteet sisältävät enemmän vaihtelua. Keskiarvomalli ennustaa saman todennäköisyyden kaikille päiville, viikonpäivämalli samalle viikonpäivälle ja logistinen regressiomalli samalle viikonpäivän ja kuukauden yhdistelmälle. Viikonpäivämallin ennusteet ovat tämän takia kuvassa viikon syklissä ja logistisen regressiomallin ennusteet tämän lisäksi myös kuukauden syklissä. Kuukauden vaihtuessa maaliskuusi logistisen regressiomallin kuvaajassa tapahtuu selkeä muutos.



Kuva 1: Junan 2091 kulkutieto ja eri malleilla ennustetut peruuntumistodennäköisyydet. Toteuman arvo 1 kuvaa junan peruuntumista, ja toteuman arvo 0 sitä, että juna ei peruuntunut.

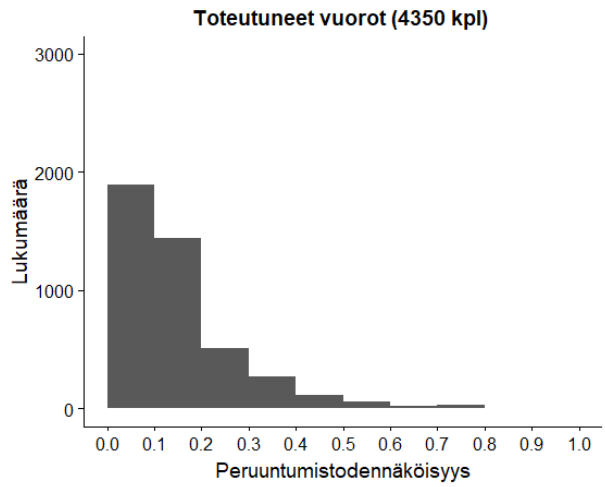
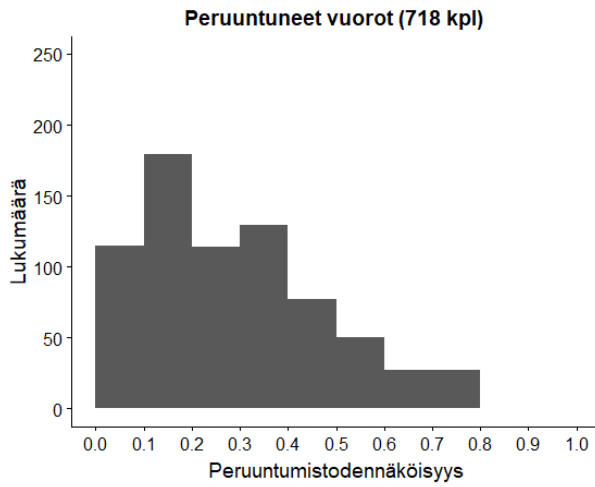
Yhdelle junalle voidaan projektiryhmämme luoman työkalun avulla piirtää myös eroteltu aikasarjakuvaaja. Tämä kuvaaja eroaa edellisestä siten, että peruuntuneet ja toteutuneet junavuorot on eroteltu erillisiin kuviin. Tämä rikkoo aika-akselin säännöllisyyden, mutta helpottaa ennusteen hyvyyden visuaalista tarkastelua. Kuvassa 2 on esimerkkinä eroteltu aikasarjakuvaaja junalle 2091 ja logistiselle regressiomallille.



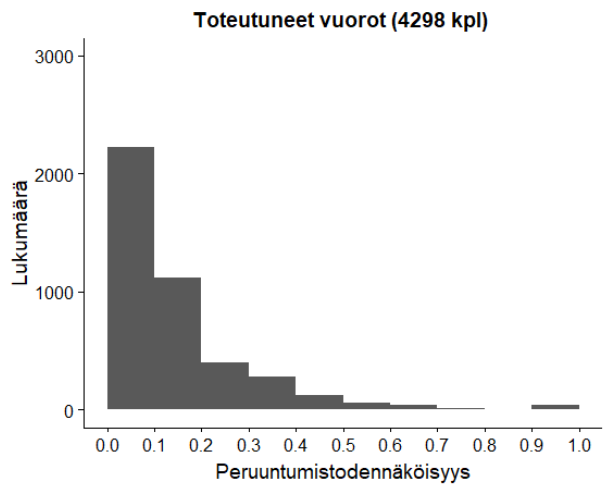
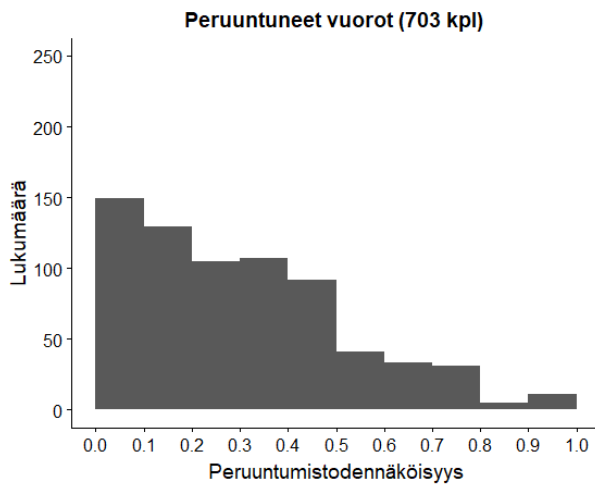
Kuva 2: Eroteltu aikasarjakuvaaja junalle 2091 ja logistiselle regressiomallille.

Kaikkien junavuoroille ennustettujen peruuntumistodennäköisyyksien visualisoimiseksi kerralla toteutimme työkalun funktion, jolla piirretään histogrammeja ennusteista. Peruuntumistodennäköisyyksien jakauma kuvataan histogrammilla erikseen peruuntuneille ja toteutuneille junavuoroille. Ennustemalli on sitä parempi, mitä pienempiä todennäköisyyksiä se antaa toteutuneille vuoroille ja mitä suurempia peruuntuneille vuoroille. Toteutuneiden vuorojen histogrammissa peruuntumistodennäköisyyksien pitäisi siis pakkautua vasempaan laitaan ja peruuntuneiden vuorojen tapauksessa puolestaan oikeaan laitaan. Kuvassa 3 on histogrammit kaikille kolmelle ennustemallille.

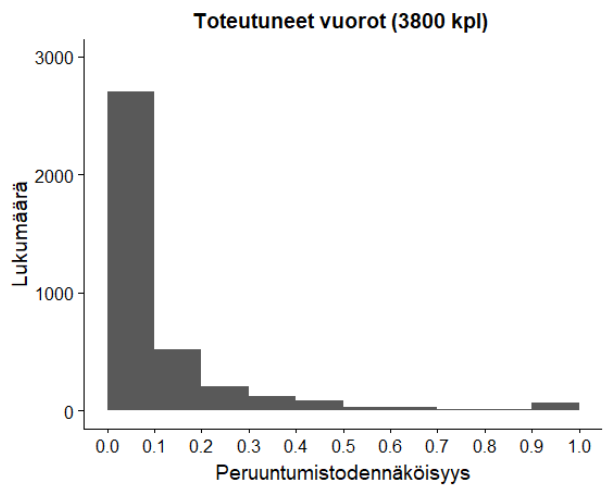
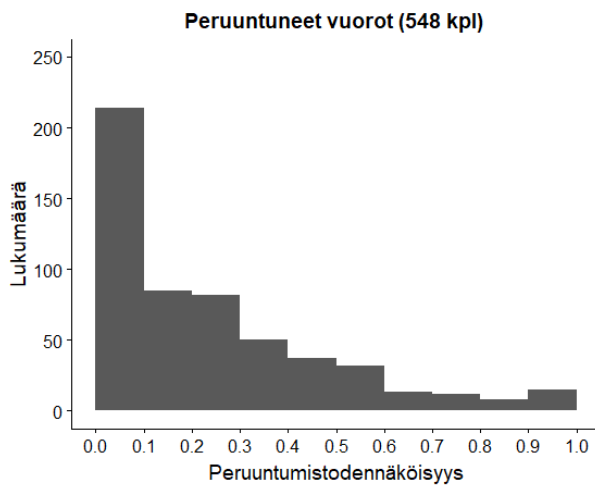
Aikavälille 2017-02-11 - 2017-03-03 KA-mallilla lasketut peruuntumistodennäköisyydet



Aikavälille 2017-02-11 - 2017-03-03 viikonpäiväKA-mallilla lasketut peruuntumistodennäköisyydet



Aikavälille 2017-02-11 - 2017-03-03 logit-mallilla lasketut peruuntumistodennäköisyydet



Kuva 3: Histogrammit kaikista kolmen eri ennustemallin tuottamista peruuntumistodennäköisyyksistä erikseen peruuntuneille ja toteutuneille junavuoroille.

Kuvasta 3 havaitaan sama ilmiö kuin taulukosta 2: Monimutkaisemmilla malleilla ei pystytä luomaan ennusteita yhtä usealle junavuorolle kuin yksinkertaisimmilla malleilla. Kävimme läpi syitä tähän kappaleessa 5.5.1. Histogrammien perusteella logistinen regressio vaikuttaa ennustavan toteutuneet vuorot muita malleja paremmin. Toisaalta se näyttäisi ennustavan peruuntuneita vuoroja hieman muita malleja heikommin, mutta niitä on myös huomattavasti vähemmän.

6 Käyttöliittymä

Projektin käyttöliittymänä on R-ohjelmointikielellä toteutettu ohjelma, joka tuottaa päätöksentekoa tukevan taulukon resurssisuunnittelijoille. Käyttäjän tulee itse syöttää ohjelmalle tarkasteltavan kustannuspaikan numero (esimerkiksi 617 Kouvolalle), tarkasteltavan ajanjakson alku- ja loppupäivämäärä sekä nimi tiedostolle, josta löytyy tiedot kuljettajien työvuoroista ja niiden työtehtävistä suunnittelujaksolle. Taulukkoa lukemalla suunnittelija pystyy valitsemaan työvuorot työvuoroluetteloon kustannustehokkaasti. Käyttöliittymä jättää suunnittelijalle kuitenkin valitsemiseen harkintavaraa päätöksenteko-ongelman monimutkaisuuden takia.

Tärkeimpiä ominaisuuksia käyttöliittymälle olivat helppokäyttöisyys, ymmärrettävyys, mahdollisuus vaivattomaan jatkokehittämiseen sekä työkalun rajoitteiden selkeys. Päätöksenteko-ongelmasta pyrittiinkin löytämään tärkeimmät ulottuvuudet, jotka pystyttäisiin mallintamaan hyvin. Lähestymistavaksi valittiin työvuorojen tarkastelu ns. makrotasolla. Sen sijaan, että tutkittaisiin yksitellen jokaisen ajettavan junan peruuntumistodennäköisyyttä ja sen vaikutusta muihin työtehtäviin, päätettiin tarkastella työtehtävien muodostamia työvuoroja kokonaisuuksina. Käyttöliittymän teko jakaantui työvuoroon liittyvän peruuntumistodennäköisyyden muodostamiseen, peruuntumisten vaikutusten mallintamiseen ja näiden tietojen yhdistämiseen päätöksenteon avuksi.

6.1 Peruuntumistodennäköisyys työvuorolle

Ennustusmallit-kappaleessa kerrottiin, kuinka tavarajunan peruuntumistodennäköisyys voidaan ennustaa eri tavoin. Käyttöliittymän tarvitsee kuitenkin tietää, kuinka todennäköisiä peruuntumiset ovat koko työvuorossa. Yksittäisten junien peruuntumistodennäköisyydet pitää siis koota työvuoroa koskevaksi tunnusluvuksi. Tämän laskemisessa oleellista on se, että yksikin peruutusherkkä juna vaikuttaa työvuoroon paljon.

Ratkaisuksi keksittiin vertailla työvuoroja sillä todennäköisyydellä, jolla yhtäkään työvuoron sisältämistä junista ei peruta. Jokaisen junan peruuntumistodennäköisyyden sijaan pitää siis ensiksi laskea todennäköisyys sille, että kyseistä junaa ei peruta. Tämän jälkeen nämä todennäköisyydet työvuoron junille tulee kertoa keskenään. Todennäköisyysmuuttuja yhdelle työvuorolle voidaan siis esittää muodossa

$$T = \prod_{i=1}^n (1 - p_i),$$

missä p_i on yhden työvuoron junan peruuntumistodennäköisyys.

6.2 Peruuntumisten vaikutus työvuoroon

Projektin kannalta oleellista on selvittää, mitä vaikutuksia junien peruuntumisilla on työvuoron järjestämiseen. Joskus yhden junan peruuntuminen saattaa tehdä koko työvuoron tyhjäksi ja kuljettajan

näin vapaaksi muihin tehtäviin. Toisinaan taas junavuoron peruuntuessa kuljettaja joutuu kuitenkin ajamaan saman välin pelkällä veturilla. Jos työvuoron tehtävien odotusarvoinen kesto on huomattavasti alhaisempi kuin suunniteltu, työvuoroa ei kannata laittaa työvuoroluetteloon. Tällä periaatteella voidaan vähentää ylityökustannuksia.

Suunnittelijoiden kanssa käytyjen keskusteluiden ja esimerkitapausten analysoinnin perusteella työvuoroista tunnistettiin kaksi päätekijää, jotka vaikuttavat työvuoron tehtävien odotusarvoiseen keston: työvuoron sisältämät siirtymätehtävät ja ns. kylmäasemilta lähtevät junat. Muita vaikuttavia tekijöitä on esimerkiksi työvuoron sisältämät tauot, mutta vaikutusten suuruus ja suunta sekä mahdolliset poikkeukset ovat näillä tekijöillä huomattavasti epäselvempiä. Valitut kaksi ominaisuutta ovat myös sellaisia, ettei niitä ole välttämätöntä laskea jokaiselle junalle erikseen, vaan työvuoroja pystytään tarkastelemaan kokonaisuuksina. Alemman tason lähestymistavalla olisi teoriassa mahdollista saada tarkempia tuloksia, mutta työvuoron junien monimutkaiset keskinäiset vuorovaikutukset ja mahdolliset poikkeustilanteet tekevät tällöin käyttöliittymän tavoitteiden mukaisesta toteutuksesta turhan hankalaa.

Kylmäasema tarkoittaa juna-asemaa, jossa ei ole veturinkuljettajia eikä vetureita (esimerkiksi Hyvinkää). Jotta juna voi lähteä kylmäasemalta, sinne pitää ensin ajaa veturi. Mikäli työvuoroon sisältyy kylmäasemalta lähtevä juna ja tätä edeltävä, kylmäasemalle saapuva juna peruuntuu, työvuoron tehtävien kesto ei lyhene, sillä kylmäasemalle pitää joka tapauksessa ajaa veturi. Mikäli tässä tapauksessa taas kylmäasemalta lähtevä juna peruuntuu, veturi pitää joka tapauksessa ajaa asemalta pois, joten kesto ei lyhene. Molempien junien peruuntuminen voidaan olettaa hyvin harvinaiseksi. Mikäli peruuntuva juna ei ole suoraan yhteydessä kylmäasemalta lähtevään junaan, vaikutus on hyvin tapauskohtainen. Yleisesti ottaen voidaan kuitenkin ajatella, että jos työvuorossa on kylmäasemilta lähteviä junia, peruuntumisten vaikutus työvuoron keston on vähäinen. Työvuorokohtaiseksi muuttujaksi lasketaan kylmäasemilta lähtevien junien ajallinen osuus työvuorosta. Mikäli tämä ajallinen osuus on suuri, mahdollinen peruuntunut juna on hyvin todennäköisesti suoraan kytköksissä kylmäasemalta lähtevään junaan. Mitä suurempi kylmäasemamuuttujan arvo on, sitä pienempi vaikutus mahdollisella peruuntumisella siis on työvuoron tehtävien keston. Matemaattisesti kylmäasemamuuttuja voidaan ilmaista muodossa

$$K = \frac{\textit{kylmäasemilta lähtevien junien kesto työvuorossa}}{\textit{työvuoron kesto}}$$

Siirtymä työvuorossa tarkoittaa esimerkiksi sitä, että kuljettaja siirtyy asemalta toiselle autolla tai matkustajana junassa. Junan peruuntuminen työvuorossa usein poistaa myös tarpeen siirtymälle. Jos työvuorossa on paljon siirtymää, junan peruuntuminen voi siis lyhentää työvuoron tehtävien kesto huomattavasti peruuntuneen junan kesto enemmän. Työvuorokohtaiseksi muuttujaksi lasketaan kaikkien muiden työtehtävien kuin siirtymätehtävien ajallinen osuus koko työvuorosta. Jos työvuorossa ei ole ollenkaan siirtymätehtäviä, siirtymämuuttujan arvo on siis 1. Jos työvuorossa kuljettaja ainoastaan ajaa junan asemalta toiselle ja palaa matkustajana takaisin, siirtymämuuttujan arvoksi tulee noin 0,5. Mitä lähempänä arvo on nolaa, sitä suurempi keskimääräinen vaikutus peruuntumisilla on tehtävien odotusarvoiseen keston. Matemaattisesti siirtymämuuttuja voidaan ilmaista muodossa

$$S = 1 - \frac{\textit{siirtymätehtävien kesto työvuorossa}}{\textit{työvuoron kesto}}$$

6.3 Todennäköisyyksien ja vaikutusten yhdistäminen

Työvuoron odotusarvoiseen keston suhteessa suunniteltuun vaikuttaa siis kolme tekijää: peruuntumistodennäköisyys, siirtymien määrä ja kylmäasemilta lähtevät junat. Näitä kolmea tekijää voidaan kuvata aiemmin määritellyillä muuttujilla T , S ja K . Nämä muuttujat ovat kaikki rajoittuneet välille $[0, 1]$ ja käyttäytyvät saman suuntaisesti. Mitä suurempi muuttujan arvo on, sitä turvallisemmin sen perusteella työvuoron voi laittaa työvuoroluetteloon ja päinvastoin. Todennäköisyysmuuttuja T on selkeästi merkittävin muuttujista, sillä ilman peruuntumisia työvuoron tehtävien kesto pysyy varmasti suunniteltuna. Työvuoron muuttujat eivät kuitenkaan välttämättä ole samaa suuruusluokkaa keskenään, vaan yhden perusteella työvuoron voisi laittaa työvuoroluetteloon ja toisen perusteella taas ei. Jotta suunnittelija voi tehdä päätöksen, muuttujien sisältämät tiedot pitää yhdistää jotenkin.

Selkein tapa yhdistää muuttujat olisi luoda arvofunktiio V , jonka numeerisen arvon perusteella työvuorot voisi järjestää yksiselitteisesti. Työvuorot laitettaisiin työvuorolistaan arvofunktion suuruuden mukaisessa järjestyksessä. Yleisin arvofunktio on additiivinen arvofunktiio, joka on siis muuttujien painotettu summa. Muuttujat kerrotaan niiden painoarvoa kuvaavilla vakioilla. Additiivinen arvofunktiio olisi yhdelle työvuorolle muotoa

$$V(n, p) = t * T(n, p) + s * S(n, p) + k * K(n, p),$$

missä n on työvuoron numero, p on päivämäärä ja t , s ja k ovat muuttujien painokertoimia.

Projektiryhmässä kuitenkin koettiin, ettei tarvittavien painokertoimien selvittämiseksi riittävällä tarkkuudella ole kurssin puitteissa resursseja. Lisäksi pidettiin todennäköisenä, ettei additiivisen funktion oletukset pätsisi ongelmassa. Esimerkiksi jos työvuoron junat ajetaan varmasti ($T \approx 1$), siirtymä- ja kylmäasemamuuttujien arvoilla ei ole mitään väliä. Peruuntumisten vaikutuksia kuvaavien muuttujien painoarvot ovat siis jollakin tavalla todennäköisyysmuuttujan funktioita. Painoarvot ovat myös jossakin määrin tapauskohtaisia eli työvuorosta riippuvia, sillä muuttujat kuvaavat vain yleisellä tasolla tärkeimpiä tekijöitä.

Suunnittelija pystyy näkemään työvuoron ominaisuudet paremmin itse kuin yksinkertaistettu malli. Tietoa todennäköisyyksistä hänellä ei kuitenkaan ole, joten tämä on ainoa mallin tarjoama lisäinformaatio suunnittelijalle. Siispä käyttöliittymä tuottaa suunnittelijalle taulukon, jossa muuttujat on esitelty erikseen. Suunnittelijan vastuulla on kompensoida todennäköisyysmuuttujan alhaisia tai keskisuuria arvoja siirtymä- ja kylmäasemamuuttujilla. Esittämällä työvuoron pääominaisuudet toisistaan eroteltuna voidaan kuitenkin myös nopeuttaa suunnittelijan työtä, sillä hän pystyy kiinnittämään huomionsa kustannusten kannalta tärkeimpiin työvuoroihin eikä jokaisen työvuoron sisällön tutkiminen ole välttämätöntä.

6.4 Päätöksentekotaulukko

Esimerkit suunnittelijan saamasta taulukosta on esitetty kuvissa 4 ja 5. Taulukossa rivit vastaavat tietyn kustannuspaikan työvuoroja. Ensimmäisessä sarakkeessa on työvuoron numero ja toisessa sitä vastaava päivämäärä. Sama työvuoronumero voi olla eri päivinä. Yleensä työvuorot ovat tällöin muuten täysin identtiset, eli ne koostuvat samoista junista. Kolmannessa sarakkeessa on todennäköisyysmuuttuja kahden desimaalin tarkkuudella. Viimeiset sarakkeet vastaavat kylmäasema- ja siirtymämuuttujaa. Taulukko tulee järjestää todennäköisyysmuuttujan arvon perusteella. Tämä järjestys kertoo lähtökohtaisesti, mitkä työvuorot kannattaa sijoittaa työvuoroluetteloon.

Päätöksenteon helpottamiseksi kylmäasema- ja siirtymämuuttuja on muutettu jatkuvasta kategoriseksi. Muuttujien arvot jaetaan kolmeen ryhmään. Kylmäasemamuuttujan alimmassa ryhmässä muuttujan arvo on tasan nolla. Ryhmän nimenä on "EI KYLMÄASEMIA". Ylimmässä ryhmässä muuttujan arvo on suurempi kuin 0,4. Ryhmän nimenä on "PALJON KYLMÄASEMIA". Väliin jäävä ryhmä on nimeltään tyhjä. Siirtymämuuttujan alimmassa ryhmässä muuttujan arvo on alle 0,7. Ryhmän nimenä on "PALJON SIIRTYMÄÄ". Ylimmässä ryhmässä muuttujan arvo on suurempi kuin 0,9. Ryhmän nimenä on "EI SIIRTYMÄÄ". Väliin jäävä ryhmä on tässäkin tapauksessa nimeltään tyhjä. Kategorinen muuttuja on väritetty vihreällä, mikäli pelkästään sen perusteella työvuoro olisi turvallista sijoittaa työvuoroluetteloon ("PALJON KYLMÄASEMIA" ja "EI SIIRTYMÄÄ"). Oranssi väri taas tarkoittaa päinvastaista ("EI KYLMÄASEMIA" ja "PALJON SIIRTYMÄÄ").

Taulukko tulee ensin järjestää todennäköisyyden mukaan joko laskevasti tai nousevasti. Kuvassa 4 taulukko on järjestetty laskevasti. Kaikissa kuvassa näkyvissä työvuoroissa todennäköisyysmuuttujan arvo on hyvin lähellä yhtä, eli melko varmasti yhtäkään junaa ei peruta. Kylmäasemia ja siirtymiä koskevia tietoja ei tarvitse siis edes katsoa, vaan kaikki työvuorot voidaan turvallisesti sijoittaa työvuoroluetteloon. Suurin osa työvuoroista tulee sijoittaa työvuoroluetteloon heittomiesten vähäisen määrän takia, joten kylmäasema- ja siirtymämuuttujien tietoja tarvitsee ottaa huomioon vasta selvästi matalammilla todennäköisyysmuuttujan arvoilla.

Kuvassa 5 taulukko on järjestetty nousevasti. Todennäköisyysmuuttujien arvo on kaikissa työvuoroissa melko matala, eli junien peruuntuminen on hyvin todennäköistä. Lähtökohtaisesti työvuoroluetteloon jätetään laittamatta ne työvuorot, joiden todennäköisyysmuuttujan arvo on alhaisin. Kylmäasema- ja siirtymämuuttujan arvot voivat kuitenkin vaikuttaa peruuntumisten vaikutuksiin, joten nekin kannattaa huomioida. Mikäli molemmat muuttujat ovat vihreitä, työvuoron laittaminen työvuoroluetteloon voi olla turvallista, vaikka peruuntuminen olisikin todennäköistä. Toisaalta jos molemmat muuttujat ovat oransseja, peruuntumisella voi olla suuri vaikutus, vaikkei todennäköisyysmuuttujan arvo olisikaan niin alhainen. Suurimmassa osassa työvuoroista ei kuitenkaan ole siirtymää, joten tällainen on harvinaista. Useimmissa työvuoroissa molemmat muuttujat eivät ole samaa väriä, jolloin suunnittelijan pitää käyttää omaa harkintaansa tai antaa todennäköisyysmuuttujan ratkaista.

Show 25 entries

Search:

tyovuoro	paivamaara	todennakoisyys	kylmaasemat	siirtymia
30225	2017-01-31	1	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30274	2017-01-31	1		EI SIIRTYMÄÄ
10202	2017-01-31	1	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
40347	2017-01-31	1	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30093	2017-01-31	1	EI KYLMÄASEMIA	PALJON SIIRTYMÄÄ
10295	2017-02-01	1		
30093	2017-02-01	1	EI KYLMÄASEMIA	PALJON SIIRTYMÄÄ
10202	2017-02-01	1	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30225	2017-02-01	1	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30447	2017-01-31	0.98		EI SIIRTYMÄÄ
10558	2017-01-31	0.98	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30447	2017-02-01	0.98	EI KYLMÄASEMIA	
10214	2017-01-31	0.97		EI SIIRTYMÄÄ
30075	2017-02-01	0.97	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30115	2017-01-31	0.94		PALJON SIIRTYMÄÄ

Kuva 4: Laskeva järjestys todennäköisyyksille

Show 25 entries

Search:

tyovuoro	paivamaara	todennakoisyys	kylmaasemat	siirtymia
483	2017-01-31	0.07	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
483	2017-02-01	0.07	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
10187	2017-01-31	0.08	PALJON KYLMÄASEMIA	EI SIIRTYMÄÄ
10586	2017-01-31	0.14	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
10586	2017-02-01	0.14		EI SIIRTYMÄÄ
30111	2017-01-31	0.17	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30111	2017-02-01	0.17	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
30942	2017-01-31	0.18	PALJON KYLMÄASEMIA	EI SIIRTYMÄÄ
10322	2017-01-31	0.18	PALJON KYLMÄASEMIA	
30942	2017-02-01	0.18	PALJON KYLMÄASEMIA	EI SIIRTYMÄÄ
10114	2017-02-01	0.24		EI SIIRTYMÄÄ
30034	2017-01-31	0.35	EI KYLMÄASEMIA	EI SIIRTYMÄÄ
10381	2017-02-01	0.36		EI SIIRTYMÄÄ
10251	2017-01-31	0.37	EI KYLMÄASEMIA	
10251	2017-02-01	0.37	EI KYLMÄASEMIA	

Kuva 5: Nouseva järjestys todennäköisyyksille

6.5 Käyttöliittymän toiminta

Käyttöliittymänä toimii R-kielellä toteutettu skripti. Käyttäjän tulee ensin syöttää ohjelmalle kustannuspaikan numero, tarkasteltavan ajanjakson alku- ja loppupäivämäärä sekä nimi tiedostolle, josta löytyy tiedot kuljettajien työvuoroista ja niiden työtehtävistä suunnittelujaksolle. Tämän jälkeen skriptin voi suorittaa ja suunnittelija saa hetken päästä valmiin päätöksentekotaulukon. Päätöksentekotaulukko on tehty käyttäen "DT"-pakettia (RStudio 2017) ja sen voi tallentaa esimerkiksi html-muodossa. Päätöksentekotaulukkoon on tarvittaessa helppo tehdä lisäominaisuuksia.

Skripti lukee aluksi työvuorotiedot sisältävän tiedoston taulukoksi. Tiedostossa rivit ovat suunnittelujakson työtehtäviä ja sarakkeet vastaavat esimerkiksi kustannuspaikkoja, päivämääriä, junanumeroa ja tehtävän alkamis- ja loppumiskellonaikaa. Rivejä voi olla tiedostossa yli 50 000. Suurinta osaa sarakkeista ei tarvita, joten vain tarpeelliset luetaan. Tämän jälkeen skripti käy läpi luodun taulukon päivä kerrallaan. Taulukosta tehdään uudet huomattavasti pienemmät päiväkohtaiset taulukot, jossa on vain tietyn päivän ja halutun kustannuspaikan työtehtävät. Päiväkohtaisissa taulukoissa työvuoronumero vastaa vain yhtä työvuoroa.

Jokaisen työvuoron työtehtävät käydään läpi päiväkohtaisesta taulukosta ja näistä lasketaan kolme vektoria, joiden pituus on sama kuin taulukon rivien määrä. Ensimmäisessä on tieto siitä, onko työtehtävän lähtöasema miehistöpaikka vai kylmäasema (0 tai 1). Toisessa on tieto siitä, voidaanko työtehtävää pitää siirtymänä (0 tai 1). Kolmannessa on työtehtävän kesto minuutteina. Neljäs samanpituisen vektori lasketaan vielä erillisellä ennustefunktiolla. Tämä vektori kertoo jokaisen työtehtävän peruuntumistodennäköisyyden. Mikäli todennäköisyyttä ei voida laskea jollekin työtehtävälle (tehtävä on esimerkiksi siirtymä), vektorissa on tässä kohtaa arvo NA.

Näistä neljästä vektorista lasketaan tämän jälkeen päiväkohtaiset todennäköisyysmuuttujat (T), kylmäasemamuuttujat (K) ja siirtymämuuttujat (S) työvuoroille. K ja S muutetaan kategorisiksi. Kun kaikki päivät on käyty läpi, voidaan yhdistää valmiit päiväkohtaiset tiedot ja muodostaa päätöksentekotaulukko.

6.5.1 Peruuntumistodennäköisyyksien laskenta

Tällä hetkellä todennäköisyydet lasketaan keskiarvomallilla, mutta käyttäjä voi myös halutessaan käyttää kahta muuta projektiryhmän kehittämistä ennustemalleista. Päädyimme jättämään keskiarvomallin ohjelmaan oletusasetukseksi sen toimintavarmuuden vuoksi. Kahdessa muussa mallissa on ongelmana niiden riippuvuus tarkasta päivämäärästä. Käymme tässä kappaleessa tarkemmin läpi tämän ongelman.

Työvuoroaineisto ei sisällä tarkkaa tietoa mainittujen junien lähtöpäivistä. Aineistossa on mainittu ajotehtävän alkamisaika, mutta ilman lisätietoa on mahdotonta määrittää, alkaako ajotehtävä junan lähtöasemalta vai onko juna lähtenyt jo huomattavasti aiemmin, jopa edellisen vuorokauden puolella. Tarkan lähtöpäivätiedon puute aiheuttaa ennustemalleissa ongelmia. Ennustetta saatetaan pyrkiä luomaan viikonpäivälle, jona juna ei ole koko historiansa aikana kulkenut. Toisaalta ennustaminen voi onnistua, mutta ennuste kuvaakin edellisen päivän junaan käyttöliittymässä mainitun sijaan.

Projektiryhmän käyttöön ei saatu aineistoa, joka olisi sisältänyt puuttuvat aikataulutiedot riittävän pitkälle tulevaisuuteen eli työvuorojen suunnittelujakson ajalle. Näitä aikatauluja voisi päätellä aiempien aikataulujen perusteella, mutta päädyimme siihen tulokseen, että tämä ei ole järkevää. Työvuorot on koottu jonkin aikataulun perusteella, joten ongelma ei ole siinä, että tätä tietoa ei olisi olemassa. Sitä ei vain tällä aikataululla saatu projektiryhmän käyttöön. Kun tämä tieto on saatavilla, sen lisääminen

työkaluun on huomattavasti helpompaa ja laskennallisesti tehokkaampaa kuin saman tiedon kokoaminen muualta.

7 Yhteenveto

Tämän projektin tavoitteena on ollut tuottaa päätöksentekotyökalu, jonka avulla VR:n työvuorosuunnittelijat voisivat ottaa junien peruuntumisherkkyden huomioon nykyistä paremmin. Työkalun on tarkoitus auttaa työvuorosuunnittelijoita päättämään, millaiset työvuorot tulisi kiinnittää työvuoroluetteloihin ja mitkä jätetään joustavampien ”heittomiesten” hoidettavaksi niiden sisältämien peruuntumisherkkien junien vuoksi. Alkuperäisenä tavoitteena oli jakaa työkalun avulla työvuorot kolmeen kategoriaan:

- a) ’tulee sijoittaa työvuoroluetteloihin’,
- b) ’tulisi ehkä sijoittaa luetteloihin’ ja
- c) ’ei ainakaan sijoiteta luetteloihin’.

Toteutimme työkalun, joka järjestää työvuorot listaksi työvuoron sisältämien junien peruuntumisherkkyden mukaan. Järjestämisperusteena on se todennäköisyys, jolla yhtäkään työvuoron sisältämistä junista ei peruta. Työkaluun ei kuitenkaan lopulta toteutettu eksplisiittistä kategorisointia, koska sen ei lopulta ajateltu olevan suunnittelijoille hyödyllinen. Periaatteessa lista oltaisiin voitu katkaista kolmeen osaan, mutta tämä olisi ollut melko keinotekoisia.

Projektissa oli myös tavoitteena tuottaa peruuntumisherkkyden lisäksi helppokäyttöistä lisätietoa erityisesti ’ehkä’-kategorian työvuoroista. Tämä toteutettiin analysoimalla työvuoron koostumusta kahdella yksinkertaisella mittarilla, joiden perusteella käyttöliittymään tuodaan sanallisia ilmoituksia. Nämä ohjaavat suunnittelijoita tutustumaan tarvittaessa työvuoron koostumukseen tarkemmin. Valitsimme tämän esitystavan, sillä mittarit eivät anna riittävän kattavaa tietoa työvuoron koostumuksesta, jotta niihin voisi luottaa ilman lisätietoa.

Junien peruuntumistodennäköisyyden arvioimiseksi kehitettiin kolme ennustemallia: keskiarvomalli, viikonpäivämalli ja logistinen regressiomalli. Ohjelma rakennettiin siten, että uusien mallien kehittäminen on helppoa. Mallien hyvyyden arviointiin rakennettiin menetelmiä tämän kehitystyön helpottamiseksi. Logistinen regressiomalli osoittautui tarkimmaksi kehitetyistä malleista, mutta se tarvitsee myös eniten dataa toimiakseen hyvin. Tämä malli toimii siis ainoastaan junille, jotka ovat olleet kulussa säännöllisesti ja pitkään. Logistisen regressiomallin jatkokehittäminen on houkuttelevaa, koska saimme hyviä tuloksia jo kahdella muuttujalla (viikonpäivä ja kuukausi). Kokeilemalla useampia muuttujia voitaisiin päästä vielä huomattavasti parempiin tuloksiin.

Projektiryhmän käytössä olleen aineiston vajavaisuuden takia ohjelmaan jätettiin oletusasetukseksi keskiarvomalli. Käytettävissä ollut työvuoroaineisto ei sisällä tarkkaa tietoa junien lähtöpäivämääristä, ja tätä aikataulutietoa ei saatu projekti aikataulun puitteissa ryhmän käyttöön. Tämän epävarmuuden vuoksi viikonpäivää selittävänä muuttujana käyttävät mallit saattavat antaa virheellistä tietoa, kunnes tarvittava aikataulutieto saadaan tuotua ohjelmaan.

Lähdeluettelo

- Abbink, E., Fischetti, M., Kroon, L., Timmer, G. & Vromans, M., 2005. Reinventing crew scheduling at Netherlands Railways. *Interfaces*, 35(5), ss.393-401.
- Alaeddini, A., Yang, K., Reeves, P. & Reddy, C.K., 2015. A hybrid prediction model for no-shows and cancellations of outpatient appointments. *IIE Transactions on Healthcare Systems Engineering*, 5(1), ss.14-32.
- Alfieri, A., Kroon, L. & van de Velde, S., 2007. Personnel scheduling in a complex logistic system: a railway application case. *Journal of Intelligent Manufacturing*, 18(2), s.223.
- Belton, V., 1986. A comparison of the analytic hierarchy process & a simple multi-attribute value function. *European Journal of Operational Research*, 26(1), s.7-21.
- Borndörfer, R., Sagnol, G., Schlechte, T. & Swarat, E., 2016. Optimal duty rostering for toll enforcement inspectors. *Annals of Operations Research*, ss.1-24.
- Choi, B., 2012. ARMA model identification. Springer Science & Business Media.
- CRAN. 2017. cran.r-project.org. [Viitattu: 11.5.2017] Saatavilla: <https://cran.r-project.org/web/packages/glarma/index.html>.
- Fügener, A., Brunner, J.O. & Podtschaske, A., 2015. Duty and workstation rostering considering preferences and fairness: a case study at a department of anaesthesiology. *International Journal of Production Research*, 53(24), ss.7465-7487.
- Hanne, T., Dornberger, R. & Frey, L., 2009, May. Multiobjective and preference-based decision support for rail crew rostering. In *Evolutionary Computation, 2009. CEC'09. IEEE Congress on* (ss. 990-996). IEEE.
- Huang, H.C., Chang, A.Y. & Ho, C.C., 2013. Using Artificial Neural Networks to Establish a Customer-cancellation Prediction Model. *Przeglad Elektrotechniczny*, 89(1b), ss.178-180.
- Iliescu, D.C., 2008. Customer based time-to-event models for cancellation behavior: A revenue management integrated approach. Georgia Institute of Technology.
- James, G. & Witten, D. & Hastie, T. & Tibshirani, R. 2013. *An Introduction to Statistical Learning*. 6. painos. New York: Springer Science+Business Media. 426 s. ISBN 978-1-4614-7137-0.
- Lezaun, M., Pérez, G. & Sáinz De La Maza, E., 2007. Rostering in a rail passenger carrier. *Journal of Scheduling*, 10(4), ss.245-254.
- Liikennevirasto. 2017. rata.digitraffic.fi. [Viitattu 5.5.2017]. Saatavilla: <https://rata.digitraffic.fi/api/v1/doc/index.html>.
- RStudio. 2017. rstudio.github.io. [Viitattu 11.5.2017] Saatavilla: <https://rstudio.github.io/DT/>.
- Torrance, G. W., Boyle, M. H., & Horwood, S. P., 1982. Application of multi-attribute utility theory to measure social preferences for health states. *Operations Research*, 30(6), ss.1043-1069.
- Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E. & De Boeck, L., 2013. Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3), ss.367-385.

Liite: Itsearviointi

Projektin kulku ja työmäärä

Projektityön edistäminen rakentui viikoittaisten tiimitapaamisten ympärille. Näiden lisäksi työtä rytmittivät asiakstapaamiset noin kerran kuukaudessa sekä kurssitapaamiset. Useita tiimitapaamisia hoidettiin Skypen välityksellä, erityisesti silloin kun jaetuissa työtehtävissä ei ollut tapahtunut merkittäviä käännteitä edellisen tapaamisen jälkeen. Työn rytmitys koettiin toimivaksi, ja projekti eteni tasaisesti. Loppuraportin kirjoittaminen tosin aiheutti hieman työmäärän kasautumista projektin loppupuolelle. Etäkokoustaminen helpotti kaukana toisistaan asuvien projektiryhmäläisten arkea, ja se koettiin toimivaksi tavaksi vaihtaa kuulumisia säännöllisesti. Koimme kurssin ekskursion kuormittaviksi niistä saatuun hyötyyn nähden. Erityisesti päivätöihin yhdistettynä pitkät ja pakolliset tapaamiset keskellä päivää aiheuttavat haasteita, ja usein ekskursion anti jäi melko ohueksi. Ekskursion siirtäminen aamuun tai myöhemmälle iltapäivään auttaisi, sillä tässä tapauksessa ne eivät matkustuksineen pirstaloisi työpäivää.

Työnajako koettiin toimivaksi. Johanna keskittyi peruuntumisdatan käsittelyyn ja ennustemallien luomiseen, Juhani käyttöliittymään ja Tapio ennustemallien hyvyyden arviointimenetelmiin. Projektin edetessä ryhmän jäsenille muodostui selkeät osaamisalueet, jotka auttoivat tehokkaassa työskentelyssä. Erikoistuminen aiheutti toisinaan haasteita työn yhteensovittamisessa, mutta mielestämme tehokas työnjako oli silti kannattavaa.

Koimme kurssin työmäärän vastaavan melko hyvin siitä saatavaa opintopistemäärää, mutta vain siinä tapauksessa, että projektityö sujuu mutkattomasti. Näin ei kuitenkaan ole useimmissa tapauksissa, sillä avoimet ongelmakuvaukset ja onnistumisen riippuvuus yhteistyön sujuvuudesta asiakkaan kanssa tuovat lähes aina yllättäviä haasteita. Näin kävi myös tässä tapauksessa ja kohdatut ongelmat kasvattivat työmäärän muita kursseja suuremmaksi. Kurssi tulisi mitoittaa siten, että yllättäviä haasteita oletetaan syntyvän kaikille ryhmille.

Haasteet

Projektia voidaan pitää yleisesti onnistuneena, sillä asiakas vaikuttaa olevan työkaluun tyytyväinen ja ryhmätyöskentely sujui hyvin loppuun saakka. Olemme tyytyväisiä työkalun ominaisuuksiin ja opimme muutamia uusia menetelmiä kurssin aikana. Projekti oli mielenkiintoinen ja tapaamiset asiakkaan kanssa olivat rakentavia. Projektiryhmälle muodostui kevään edetessä hyvä yhteishenki, mikä auttoi ongelmien kohtaamisessa.

Koimme haastavaksi asiakkaan sitoutuneisuuden vähenemisen projektin edetessä. Ensimmäisessä tapaamisessa olimme erittäin vaikuttuneita asiakkaan sitoutumisesta tiiviiseen kommunikaatioon, sekä lupauksesta irrottaa työntekijöiden aikaa tähän projektiin. Yksi asiakkaan puolen vastuuhenkilöistä siirtyi toisiin tehtäviin pian projektin alun jälkeen, mikä aiheutti hieman epäselvyyttä. Puolivälin jälkeen asiakkaan kiinnostus projektia kohtaan tuntui vähenevän merkittävästi, mikä aiheutti motivaation laskua myös projektiryhmässä. Toisaalta tämä oli erittäin hyvä työmäärän kannalta verrattuna tilanteeseen, jossa asiakas olisi innoissaan pyytänyt jatkuvasti lisäominaisuuksia. Halusimme kuitenkin toteuttaa ne ominaisuudet, jotka olimme alussa luvanneet. Emme paketoineet projektia, vaikka se tuntui olevan asiakkaan puolesta jo aivan riittävän hyvä. Meidän mielestämme projektista kuitenkin puuttui vielä operaatiotutkimuksen kannalta mielenkiintoisin osa eli keskiarvoa kehittyneempi ennustemalli. Tässä

kehitysvaiheessa kommunikaatio asiakkaan kanssa oli merkittävästi hitaampaa kuin alkukeväästä. Tässä vaiheessa tuki kurssin puolelta olisi ollut arvokasta, jotta ryhmällä olisi säilynyt tunne siitä, että tekemisillämme on merkitystä.

Suurimmat yksittäiset haasteet projektissa liittyivät peruuntumisen vaikutusten arviointiin sekä peruuntumistietojen ja työvuorotietojen yhteensovittamiseen. Yksittäisen junan peruuntumisten vaikutukset työvuoroon ovat hyvin moninaisia. Näiden mallintamiseen kulutettiin huomattava määrä aikaa ja energiaa, vaikka lopputulos ei ollut kovin vaikuttava. Lopulta projektiryhmä tuli siihen tulokseen, että peruuntumisten vaikutusten arvioinnissa kannattaa edelleen luottaa suunnittelijoiden asiantuntemukseen. Mallit yksinkertaistaisivat ja piilottaisivat tärkeää informaatiota aivan liikaa ollakseen kovin hyödyllisiä. Jälkikäteen ajateltuna peruuntumisen vaikutusten arviointi olisi ehkä kannattanut jättää kokonaan asiakkaan sisäiseen kehitystyöhön projektiryhmän keskittyessä peruuntumisten ennustamiseen. Nyt projektiryhmä käytti huomattavan määrän aikaa työvuorosuunnittelun yksityiskohtien ymmärtämiseen. Kurssin tavoitteiden kannalta olisi ollut hyödyllisempää käyttää tämä aika operaatiotutkimuksen menetelmien parissa, esimerkiksi syventymällä laajemmin logistisen regressiomallin teoriaan ja testaamalla sitä useammilla muuttujilla.

Asiakkaan tarjoama data aiheutti teknisiä haasteita vaadittujen aikamuunnosten vuoksi. Projektiryhmän käytössä olleet työvuorotiedot kertoivat ajotehtävän alkamisajan Suomen ajassa. Ne eivät kuitenkaan sisältäneet informaatiota siitä, milloin kyseinen juna lähtee ensimmäiseltä asemaltaan. Liikennevirastolta saatavat junien peruuntumistiedot ovat puolestaan UTC-ajassa ja merkittynä junan lähtöpäivän mukaan. Ymmärsimme vasta hyvin myöhään, että ajotehtävät voivat alkaa myös väliasemilta. Aikaleimojen saaminen vastaamaan toisiaan osoittautui liian vaikeaksi, sillä asiakas ei pystynyt, ainakaan tällä aikataululla, tuottamaan projektiryhmän käyttöön aikataulutietoja työvuorojen suunnittelujaksolle. Lopullisessa työkalussa ei tämän vuoksi voitu käyttää muita kuin keskiarvomallia ennusteiden luomiseen. Muuten joissakin tapauksissa ennustetta oltaisiin pyritty luomaan vaikkapa torstaille tilanteessa, jossa kyseinen juna ei ole koko historiansa aikana kulkenut torstaisin. Tämä ei kuitenkaan tee muita malleja käyttökelvottomiksi, sillä tilanteen korjaaminen on melko helppoa, jos työkaluun saadaan tuotua tarvittavat aikataulutiedot.

Projektiryhmän toiminnan kehittäminen

Olimme samaa mieltä siitä, että ryhmätyönä tehtävän ohjelman tuottaminen ilman siihen kehitettyjä työkaluja aiheutti haasteita. Tämän kokoinen koodausprojekti tehtäisiin palkkatyönä hyvin suurella varmuudella GitHubiin, jossa eri ihmisten kirjoittaman koodin yhteensovittaminen voitaisiin toteuttaa järkevästi. Aikaresurssien vähyyks suhteessa korkealle asetettuihin kurssitavoitteisiin kuitenkin aiheutti sen, että GitHubin opettelu päätettiin olevan toissijaista. Vain yksi projektiryhmän jäsen oli käyttänyt kyseistä palvelua ennen, joten koodin versionhallinnassa tyydyttiin käyttämään pilvikansiota. Opiskelijoiden heittäminen tuntemattoman ongelmakentän äärelle on toki kurssin idea. Järjestelyissä pitäisi kuitenkin huolehtia siitä, että uusien työkalujen ja menetelmien omaksumiseen olisi varattu aikaa.

Järjestimme koodin yhteensovittamiseksi yhden työpajan, jossa kokoonnuimme viikonloppuna koko päiväksi työstämään projektia. Siellä saimme paikattua koodin yhteensovittamisessa työkalujen puutteiden ja etätyöskentelyn paljouden aiheuttamia haasteita. Työpaja koettiin hedelmälliseksi, ja pohdimme myös, olisiko näitä pitänyt olla useampiakin kevään aikana. Yksi projektiryhmän jäsen on kuitenkin ollut kevään kokopäivätöissä, joten pidemmät tapaamiset olisi pitänyt järjestää viikonloppuisin.

Yhteisen ajan löytäminen olisi ollut haastavaa, mutta yhden tai kahden samankaltaisen työpajan järjestämiseen olisi kannattanut pyrkiä.