



Ikä Tappaa Tieliikenteessä! – Osa 3 – Ennustemalli

Jatkamme tämän artikkelin edellisessä osassa mainitun Suomen Tieliikenneonnettomuudet 2005-2017 datan käsittelyä, tällä kertaa teemme ennustemallin joka pyrkii ennustamaan todennäköisintä onnettomuusluokkaa annetuilla muuttujilla.

[Väylävirasto](#) on julkaissut tietoaaineiston [Tieliikenneonnettomuudet](#) lisenssillä [Creative Commons Attribution 4.0 International License](#). Tietoaaineistossa on kullekin vuodelle kolme toisiinsa linkittyvää tiedostoa. Onnettomuudet-tiedostossa on onnettomuudet, joilla kullakin yksilöivä tunnus, ja massiivinen määrä taustatietoa itse onnettomuustilanteen tekijöistä. Henkilöt-tiedostossa on henkilötietoa, joilla kullakin yksilöivä tunnus ja se linkittyy onnettomuustunnuksella onnettomuudet-tiedostoon. Osalliset-tiedosto linkittyy sekä onnettomuustunnuksen että henkilötunnuksen kautta edellisiin tiedostoihin, tämä tiedosto kertoo lähinnä kunkin osallisen ajoneuvotiedon, sekä loukkaantumis/kuolemis tiedon, joka löytyy myös henkilöt-tiedostosta. Tietoaaineiston tehokas käyttö vaatii jonkin verran relaatiotietokanta tietämystä, että voi yhdistää kunkin tiedoston sisältämän datan toisiinsa.

Muuttujat

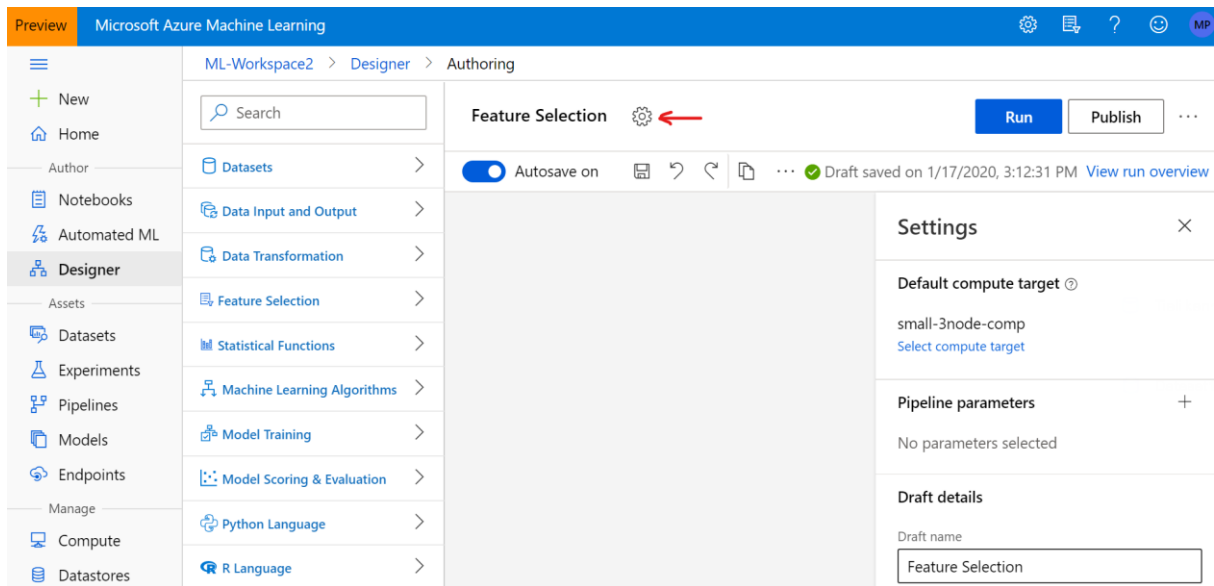
Palaamme hiukan myöhemmin syihin miksi esim. onnettomuusluokkia on yhdistelty nimelle “UusiOnLk” ja millä logiikalla, sekä miksi jokin onnettomuusluokka tai sääluokka on jopa poistettu tietoaaineistosta. Ennustemallin luonti on iteroivaa toimintaa, joten siinä pompitaan kaikkien tässä artikkelisarjassa esitettyjen vaiheiden välillä, niiden esittäminen täysin kronologisessa järjestyksessä loisi itse artikkelista sekavan, joten uusien tietoaaineistojen luonti ja käsittely on esitetty rinnakkain jotta niiden eroja päästään paremmin vertailemaan. Todellisuudessa yksi tietoaaineisto käsiteltäisiin kerrallaan alusta loppuun, jonka jälkeen havaitaan iterointitarpeet ja luodaan siltä pohjalta uusi tietoaaineisto.

Edellisessä osassa loimme Access-tietokannan nimeltä “Tieliikenneonnettomuudet.accdb” ja sinne kaksi kyselyä nimeltään “masterQuery” ja “datasetQuery”. Mene “datasetQuery”:n Design näkymään ja poista valinta Show “UusiOnLk” sarakkeesta – Save - Run. Viedään kumpikin kyselyistä tekstitiedostoon ja nimetään ne “master1.csv” ja “datasetOnLk1.csv”, valitse “Include Field Names on First Row” ja “Text Qualifier”: None. Kun vienti tehty, mene “datasetQuery”:n Design näkymään, poista valinta Show “Onluokka” sarakkeesta ja lisää valinta Show “UusiOnLk” sarakkeeseen – Save – Run, jonka jälkeen vie tekstitiedostoon nimeltä “datasetUusiOnLk1.csv”.

Tämän jälkeen kirjaudutaan edellisessä osassa luotuun [Azure Machine Learning työtilaan](#) (Preview:ssa kirjoituksen hetkellä) ja tuodaan tiedostot Datasets osion kautta tietoaaineistoiksi.

Huom! Preview vaiheessa olevilla palveluilla voi olla käyttökatkoja, muutoksia ja ongelmia, joten niitä ei missään nimessä kannata käyttää kokeilua vakavampaan tekemiseen. Kirjoittamisen hetkellä oli joitakin ongelmia isoimman tiedoston kohdalla tunnistaa sarakeotsikoita, jonka sai kierrettyä lataamalla tiedosto manuaalisesti työtilan blob-varastoon käyttäen Azure Storage Explorer ohjelmistoa ja rekisteröimällä työtilassa tiedosto Dataset:iksi.

Kun tietoaaineistot ladattu, kokeillaan löytää parhaimmat muuttujat (feature selection) “master1” tietoaaineistosta. Mennään ensin työtilan Designer osioon ja luodaan uusi Pipeline, anna nimi “Feature Selection” ja asetuksissa aseta käytettävä palvelinklusteri (Compute – Training Clusters).



Pudotetaan kanvaasille (vertikaalisesti päällekkäin tässä järjestyksessä) “master1” tietoaaineisto (Datasets), Select Columns in Dataset ja Edit Metadata (Data Transformation), sekä Filter Based Feature Selection (Feature Selection) ja yhdistetään kunkin osan Output-portti seuraavan Input-porttiin.

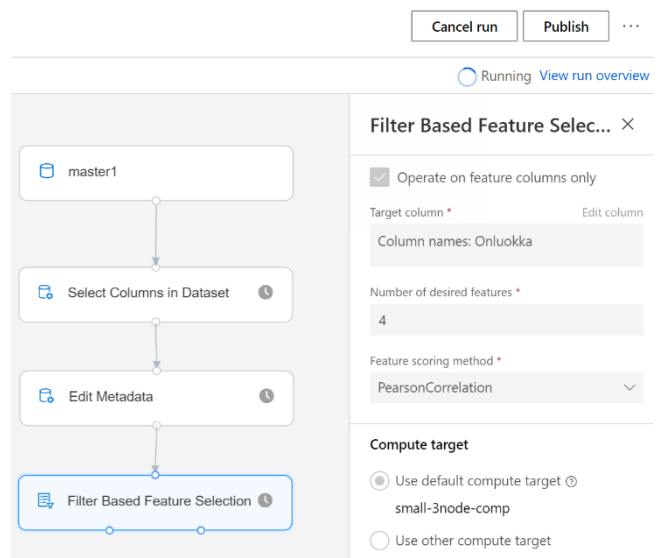
Seuraavaksi klikkaa Select Columns in Dataset ja valitse seuraavat sarakkeet:

- Tie
- Kk
- Tunti
- Vkp
- Sää
- Lämpötila
- Ikä
- Sukupuoli
- Onluokka

Sen jälkeen klikkaa Edit Metadata ja valitse Column osioon All Columns, sekä Fields kohtaan Features muiden ollessa Unchanged valinnalla.

Lopuksi klikkaa Filter Based Feature Selection, valitse Operate on feature columns only, valitse Target Column osioon Column names: Onluokka, Number of desired features: 4, sekä Feature scoring method: PearsonCorrelation.

Save – Run. Luo uusi kokeilu avautuvassa Set up pipeline run dialogissa.



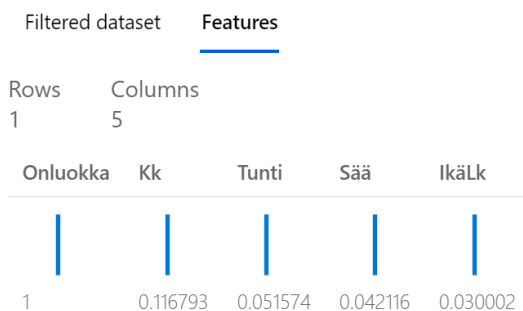
Kun ajo suoritettu, oikeaklikkaa Filter Based Feature Selection – Visualize Features. Tämän mukaan ikä ei olisi merkittävä, mutta tulee huomata että master tietoaaineistossa ei ole ikäluokka-ryhmittelyä vaan kunkin henkilön yksittäinen ikä, joita tietoaaineistossa on välillä 0-132 vuotta, joten kunkin iän määrä aineistossa on vähäinen eikä näin nouse merkittäväksi tekijäksi. Muutoin tulos tukee jo alustaviin tietoaaineistoihin valitsemiamme muuttujia (“yllättäen”, köh!).

Filter Based Feature Selection result visualization

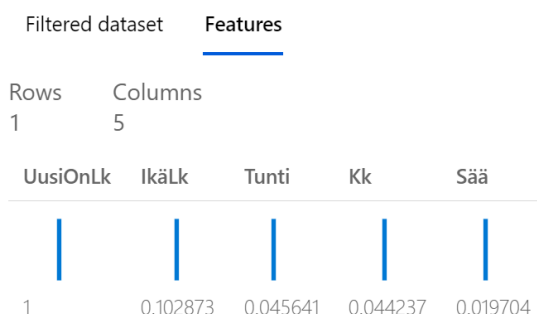


Ajetaan vastaavat muuttujavalinnat myös “OnLk” ja “UusiOnLk” tietoaaineistoja vasten valiten kaikki sarakkeet. “UusiOnLk” tietoaaineiston kanssa “IkäLk” näyttäisi nousevan merkittäväksi tekijäksi.

Filter Based Feature Selection result visualization



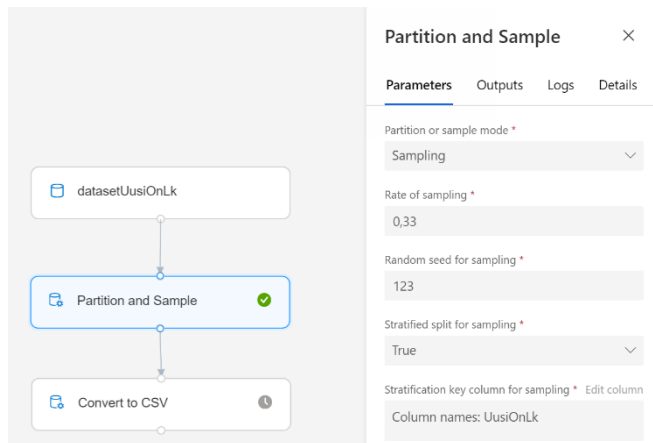
Filter Based Feature Selection result visualization



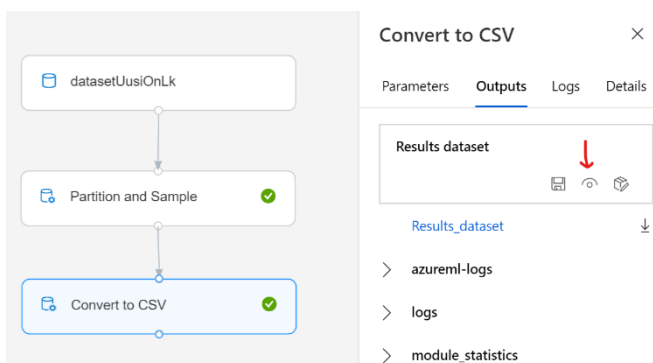
Tietoaaineiston tasapaino

Koska “UusiOnLk” tietoaaineistossakin on vielä epätasapainoa onnettomuusluokkien välillä, esim. jalakulkijaonnettomuudet aliedustettuna jopa suhteessa 1:10 muihin luokkiin verrattuna, pyritään epätasapainoa hiukan oikaisemaan tekemällä koko tietoaaineistoon aliotantaa luokkiin jotka yliedustettuna (Sampling, Stratified split).

Luodaan Designerissa uusi Pipeline, nimeltään “Feature Engineering”. Lisätään tietoaaineistoksi “datasetUusiOnLk” (Datasets), datan muokkaajaksi Partition and Sample ja muutetaan tuloste tekstitiedostoksi modulilla Convert to CSV (Data Transformation). Yhdistä tietoaaineiston Output-portti Partition and Sample Input-porttiin. Klikkaa Partition and Sample ja aseta Partition and sample mode: Sampling, Rate of sampling: 0,33, Random seed:123, Stratified split for sampling: True, Stratification key column for sampling: Column names: UusiOnLk. Yhdistä Partition and Sample Output-portti Convert to CSV Input-porttiin. Save – Run – valitse “Tieliikenneonnettomuudet” kokeilu.

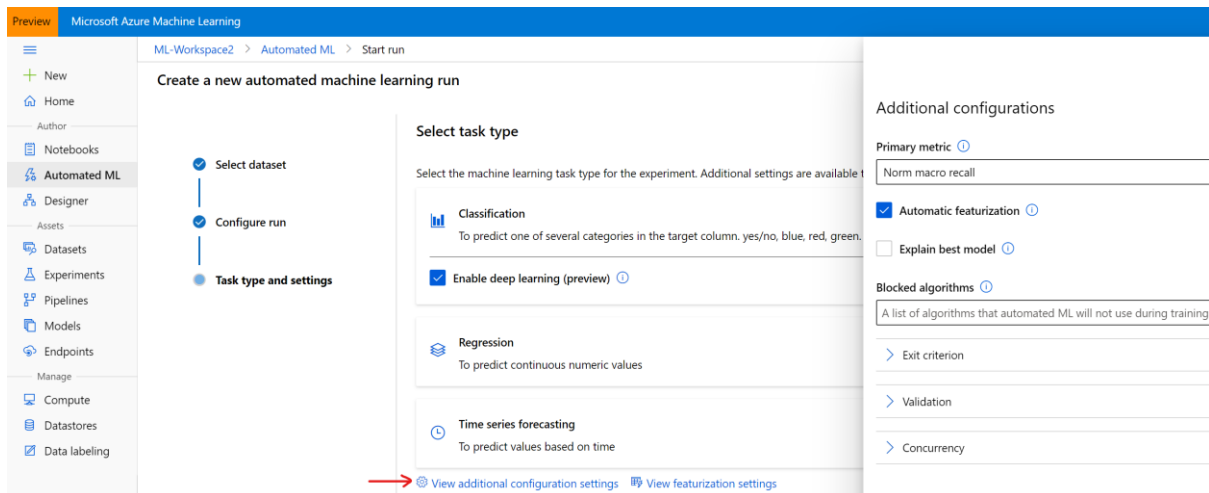


Kun ajo suoritettu, klikkaa Convert to CSV – oikealta valitaan näkymäksi Outputs – Results dataset, View Output. Tältä osin palvelu on vielä raakatimantti tai jopa kasa hiiltä, koska antaa vain polun blob-varastoon missä tietoaaineisto on. Suosittelen käyttämään Azure Storage Explorer ohjelmistoa tiedoston hakuun. Register dataset vaihtoehto ei luo Automated ML yhteensopivaa tietoaaineistoa (palautetta annettu tuoteryhmälle). Kun tietoaaineisto haettu, nimeä se “datasetUusiOnLk1-stratified.csv” ja lataa työtilaan Datasets osion kautta.



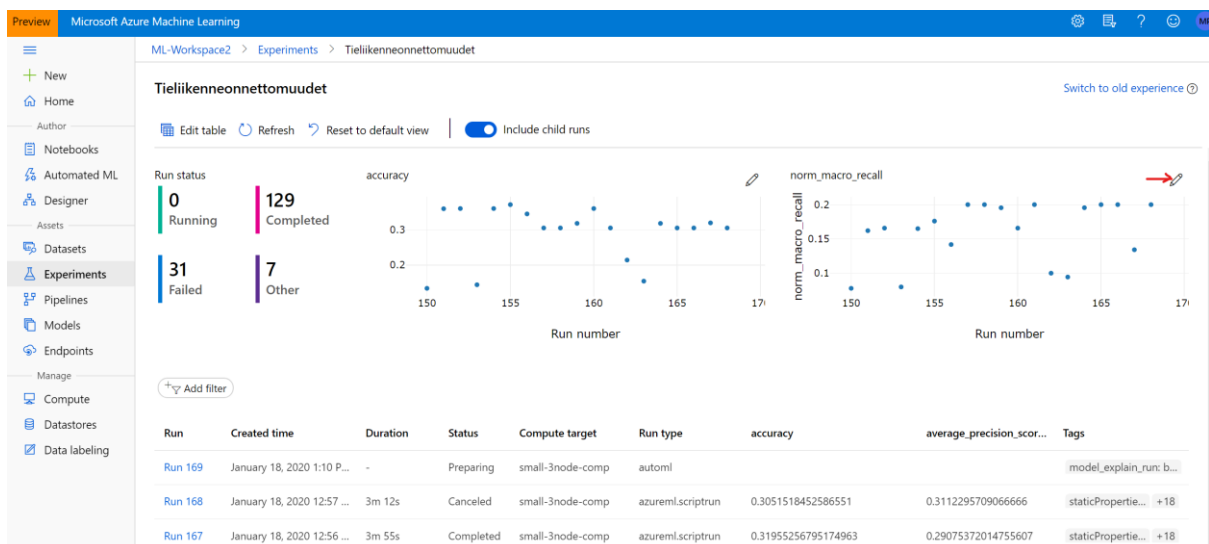
Ensimmäinen Ennustemalli

Seuraavaksi voimmekin aloittaa ennustemallien koulutuksen. Valitse vasemmalta Automated ML – New Automated ML run. Valitaan ensimmäiselle kokeilulle vähiten muokattu tietoaaineisto “datasetOnLk” – Experiment name: “Tieliikenneonnettomuudet”, Target column: “OnLuokka” – Next - valitaan Classification tyyppinen malli ja ruksataan Enable deep learning, klikataan View additional configuration settings ja asetetaan Primary metric: Norm macro recall, muut oletuksilla – Save. Valittuna oleva Automatic featurization mm. prosessoi datan vääristymät aina 1:20 suhteeseen saakka, joten haluamme verrata sen suorituskkyä juuri käsin prosessoimaamme tietoaaineistoon verrattuna.



Voit tarkistaa View featurization settings näkymässä onko esim. oikeat sarakkeet ja tarvittaessa muuttaa joitakin asetuksia. Finish. Sitten voitkin mennä lounaalle/lenkille/saunaan/muihin aktiviteetteihin noin tunniksi tai pari. Jos jonkin mallin koulutusajo on kesken ja kestänyt yli 20 minuuttia, mitä todennäköisimmin ei tule onnistumaan tai ainakaan antamaan mitään sen parempia tuloksia, joten ajon voi keskeyttää ja ko. tyyppisen mallin esim. rajata pois seuraavasta ajosta kohdassa View additional configuration settings. Esim. Support Vector Machine (SVM) jäi jumittamaan kun jossain tietoaaineistossa oli livahtanut tyhjiä/täytearvoilla olevia kenttiä mukaan, sen kerran kun meni lävitse olivat tulokset sillä huonoimmasta päästä.

Tarkastellaksesi eri mallien suorituskykyä ja visualisointeja ajojen ollessa vielä kesken, siirry vasemmalta Experiments näkymään, klikkaa "Tieliikenneonnettomuudet" kokeilua, käännä valinta Include child runs päälle. Vaihda toisen graafin mittariksi seuraamasi mittari, tässä tapauksessa Norm_macro_recall, ja tarkista graafista potentiaalisimman ennustemallin ajon numero. Ennustemallin yksityiskohtiin ja visualisointeihin pääset klikkaamalla alla olevasta taulukosta ajon numeroa.



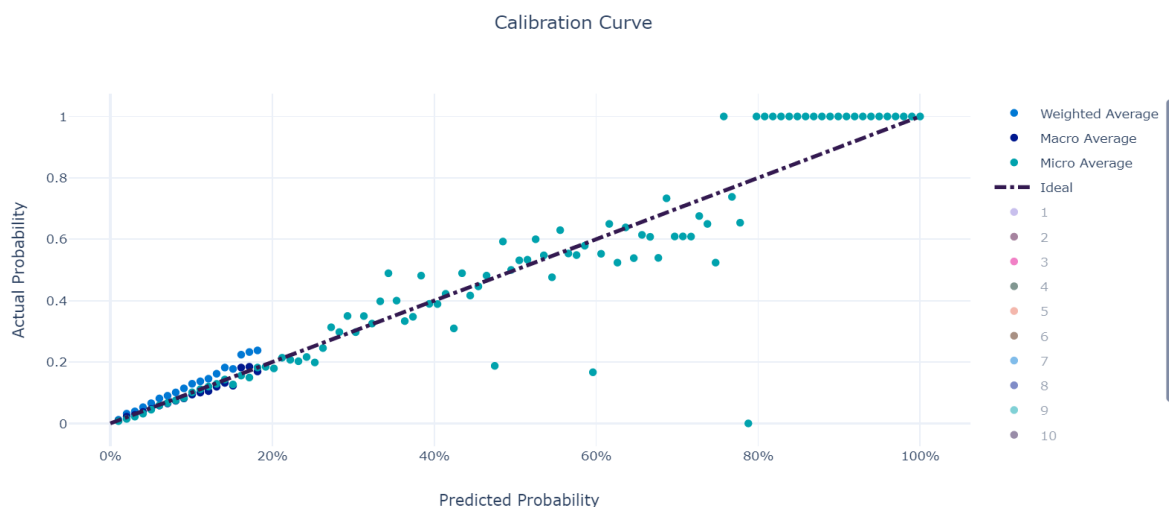
Jos haluat ajojen jo päättyttyä tarkistaa vain valitsemasi mittarin mukaan parhaimmat, käännä Experiments näkymässä Include child runs pois päältä, klikkaa ajosi numeroa alla olevasta taulukosta, vaihda Models näkymään ja selaa parhaimpien ajojen yksityiskohtia klikkaamalla kunkin algoritmin nimeä jotka menneet lävitse onnistuneesti. Data guardrails näkymästä voit tarkistaa oliko

tietoaaineistosi kuinka hyvässä kunnossa. Alla esimerkki “datasetOnLk” tietoaaineiston luokkien epätasapainosta.

Details Models Data guardrails Properties Logs Outputs

Type	Status	Description	
Class balancing detection	alerted	Classes in the training data are imbalanced.	!
Additional details			
Missing values imputation	passed	There were no missing values found in the training data.	✓
High cardinality feature detection	passed	Your inputs were analyzed, and no high cardinality features were detected.	✓

Tarkastellaanpa “datasetOnLk” tietoaaineiston parhaimman ennustemallin visuaalisia yksityiskohtia ja miten epätasapaino näkyy ennusteissa, onko Automatic featurization onnistunut korjaamaan tilannetta. Nopea vilkaisu ROC käyrään, että näyttää normaalilta. Seuraavaksi Calibration Curve, josta nähdään pysyväkö pakka kasassa kuinka kauan vai hajoaako jo alkumetreillä, eli miten seuraa ideaalilinjaa. Tällä kertaa näyttää pysyvän suhteellisen hyvin kasassa.



Seuraavaksi tarkastellaan Confusion Matrix, josta Normalized näkymä, eli kertoo ennusteen suhteellisen osumatarkkuuden kullekin oikealle luokalle (True label), numerot laskemalla vaakariviltä yhteen = 1. Ideaalitilanteessa isoimmat luvut olisivat vasemmalta ylhäältä oikealle alas menevällä lävistäjällä. Kuvasta näemme että luokat 2,3,4 ja 12 ovat aliedustettuina ennusteissa. Oikein menneissä ennusteissa yli 0,08 menevät arvot ovat satunnaista arvausta parempia tuloksia (1/12).

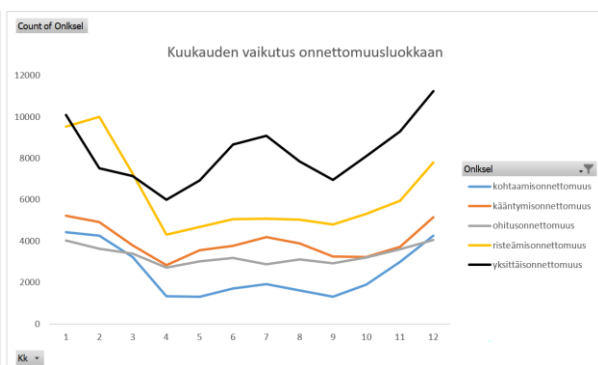
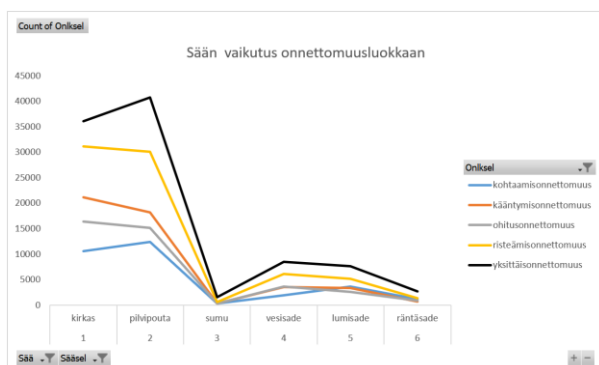


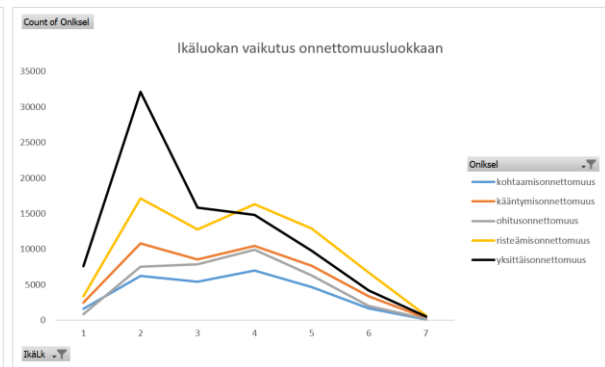
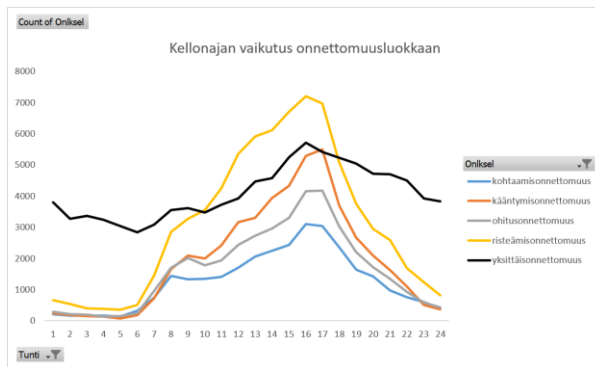
Onnettomuusluokkien yhdistely

Tietoaineiston onnettomuusluokkien jakaamaa tarkastelemalla huomaamme että luokka 12 (muu eläinonnettomuus) on huomattavan aliedustettuna (0,60%), joten siihen nähden ennustustarkkuus (12,89%) on yllättävänkin hyvä. Automatic featurization on sen osalta tehnyt hyvää työtä. Voimme taputella myös itseämme selkään, muuttujavalikoima on ollut ilmeisen onnistunut kun on pystynyt nostamaan noinkin aliedustetun luokan kohtalaiselle ennustetarkkuudelle. Lisäksi huomattavaa on, että luokan 12 kohdalla ennusteita on mennyt luokkien 10 (hirvi) ja 11 (peura) puolelle, eli luokkien yhdistäminen on varteenotettava vaihtoehto jo eläimellisen luonteensa puolesta.

Luokkien 2 (kääntymis), 3 (ohitus) ja 4 (risteämis) osalta on tutkittava vaihtoehtoja tarkemmin, mikä niissä aiheuttaa epätarkkuutta vaikka niiden osuus tietoaineistosta on hyvä ja luokassa 4 jopa erinomainen. Palatkaamme artikkelisarjan ensimmäisen osan raakadatan graafeihin.

Row Labels	Count of Onluokka	Count of Onluokka %
1 yksittäisonnettomuus	98900	17,00 %
2 kääntymisonnettomuus	47554	8,18 %
3 ohitusonnettomuus	39786	6,84 %
4 risteämissonnettomuus	74851	12,87 %
5 kohtaamissonnettomuus	30263	5,20 %
6 peräänajo-onnettomuus	78321	13,47 %
7 mopedionnettomuus	23489	4,04 %
8 polkupyöräonnettomuus	22856	3,93 %
9 jalankulkijaonnettomuus	11988	2,06 %
10 hirvionnettomuus	27584	4,74 %
11 peuraonnettomuus	44865	7,71 %
12 muu eläinonnettomuus	3480	0,60 %
13 muu onnettomuus	77684	13,36 %
Grand Total	581621	100,00 %





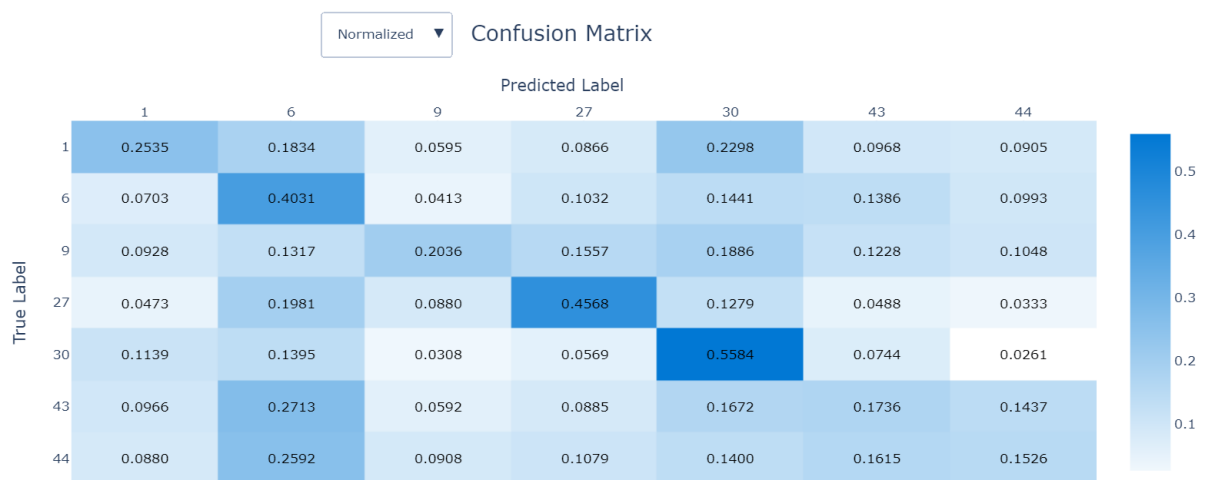
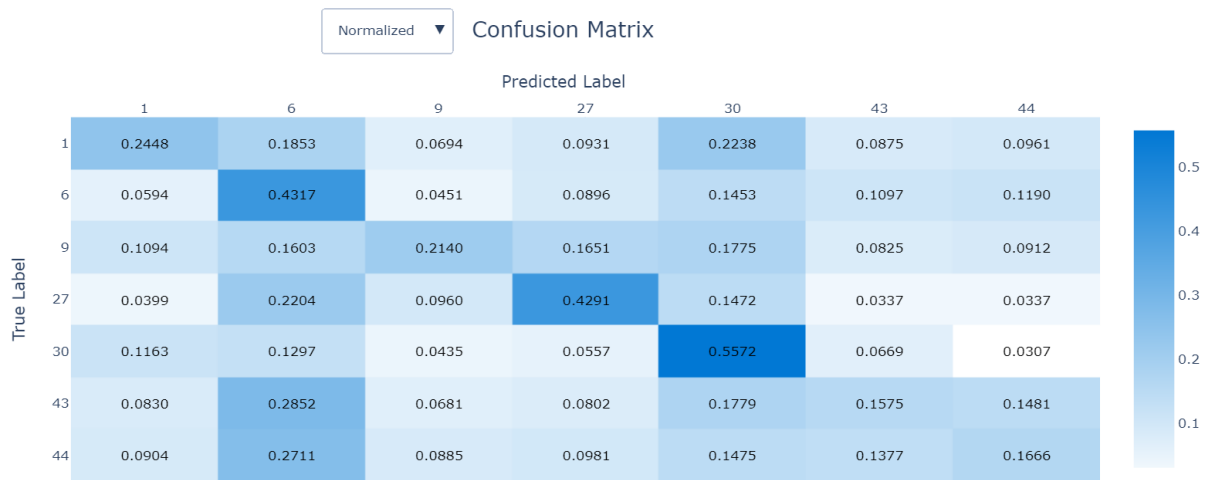
Yllä olevista kuvista voimme havaita että onnettomuusluokat 2, 3, 4 ja 5 eivät juuri profiililtaan eroa toisistaan, verrokiksi otettu onnettomuusluokka 1 (yksittäisonnettomuus), joka on musta käyrä. Ihmekään että onnettomuusluokat 2,3 ja 4 valuvat onnettomuusluokan 5 ennusteisiin. Kokeilkaamme siis onnettomuusluokkien yhdistelemistä huomioiden onnettomuusprofiilit ja suhteelliset osuudet niin että luokat olisivat tasapainossa. Lisäksi poistamme onnettomuusluokista numeron 13 (muu onnettomuus) joka on ollut onnettomuuksien kaatopaikka eikä sillä ollut erottuvaa profiilia, ensimmäisissä ennustekokeiluissa se osoittautui vaikeasti ennustettavaksi. Lisäksi säästä on poistettu sääluokka 7 (raesade) joka on erittäin harvinainen tapahtuma eikä juuri näy onnettomuustapahtumissa.

Kokeillaan seuraavanlaista onnettomuusluokkien yhdistelyä (kuva), tietoaaineisto on nimeltään "datasetUusiOnLk". Koska tämä yhdistely ei ole ensimmäinen kokeilu, on onnettomuusluokkien numerointi kasvanut uusille kymmenlukuille. Mukaan otettu vain kuljettajat ja jalankulkijat, harvemmin matkustajat ovat syyllisiä onnettomuuteen. Kuten saatamme huomata, uudet onnettomuusluokat ovat kooltaan melko lähekkäisiä, paitsi luokka 9 (jalankulkijaonnettomuus) jota ei voinut oikein yhdistää mihinkään muuhun. Seuraavaksi voimmekin seurata vaikuttaako tämä kyseisen onnettomuusluokka 9 ennustettavuuteen ja miten yhdistettyjen luokkien ennustettavuuden kävikään.

Kulj_matk (Multiple Items)		
Row Labels	Count of UusiOnLk	Count of UusiOnLk %
1		
1 yksittäisonnettomuus	71025	17,31 %
6		
6 peräänajo-onnettomuus	64999	15,84 %
9		
9 jalankulkijaonnettomuus	11163	2,72 %
27		
7 mopedi- tai polkupyöräonnettomuus	20212	4,92 %
8	21567	5,26 %
30		
10 hirvi-, peura- tai muu eläinonnettomuus	21721	5,29 %
11	40901	9,97 %
12	2904	0,71 %
43		
3 ohitus- tai kohtausonnettomuus	33314	8,12 %
5	22861	5,57 %
44		
2 kääntymis- tai risteymis- tai ohitusonnettomuus	38270	9,33 %
4	61460	14,98 %
Grand Total	410397	100,00 %

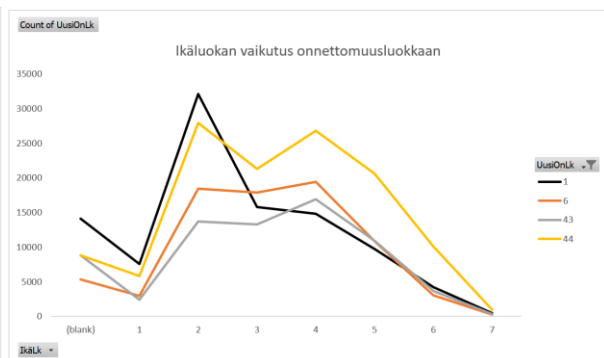
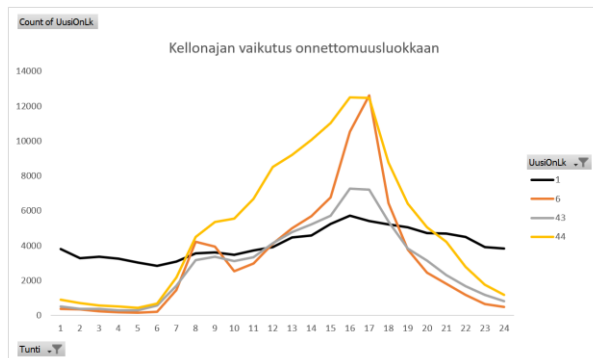
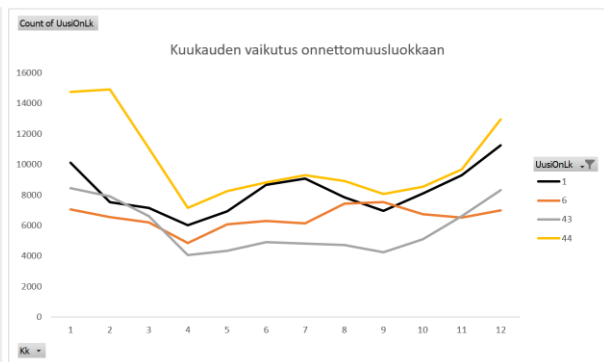
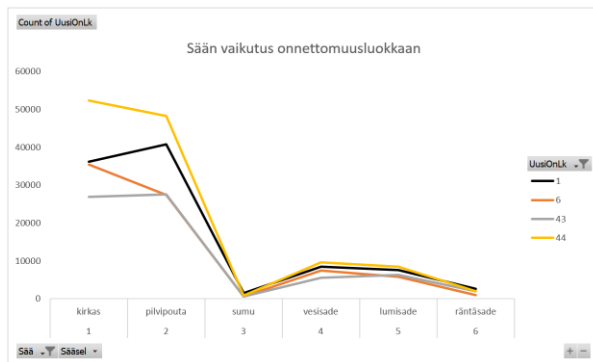
Iteroidut Ennustemallit

Tehdään ennusteajot Automated ML osiossa kuten edellä, ensin tietoaaineistolle "datasetUusiOnLk" ja kun se on valmis niin tietoaaineistolle "datasetUusiOnLk-stratified". Alla verrattu kummankin parhaimpien ajojen Confusion Matrixia. Oikein menneissä ennusteissa yli 0,14 menevät arvot ovat satunnaista arvausta parempia tuloksia (1/7).



Kuten huomata saattaa, erot ovat melko pieniä, eli Automatic featurization tekee hyvää työtä tasapainottaessaan eri muuttujia, ei juuri maksa vaivaa lähteä käsin tekemään samaa. Ennusteiden epätasapainot johtuvat enemmän luokkien samankaltaisista onnettomuusprofiileista. Onnettomuusluokka 6 (peräänajo) näyttäisi vievän luokilta 43 ja 44 ennusteita. Mikä erityisen ilahduttavaa, luokka 30 (hirvi, peura, muut eläimet) on parantanut huomattavasti ennustettavuuttaan. Muidenkin yhdistettyjen luokkien ennustettavuus on parempi kuin niiden erillisten osien keskiarvo olisi. Jopa piskuinen luokka 9 (jalankulkija) 2,72% tietoaaineisto-osuudellaan pitää hyvin pintansa ennustettavuudessaan.

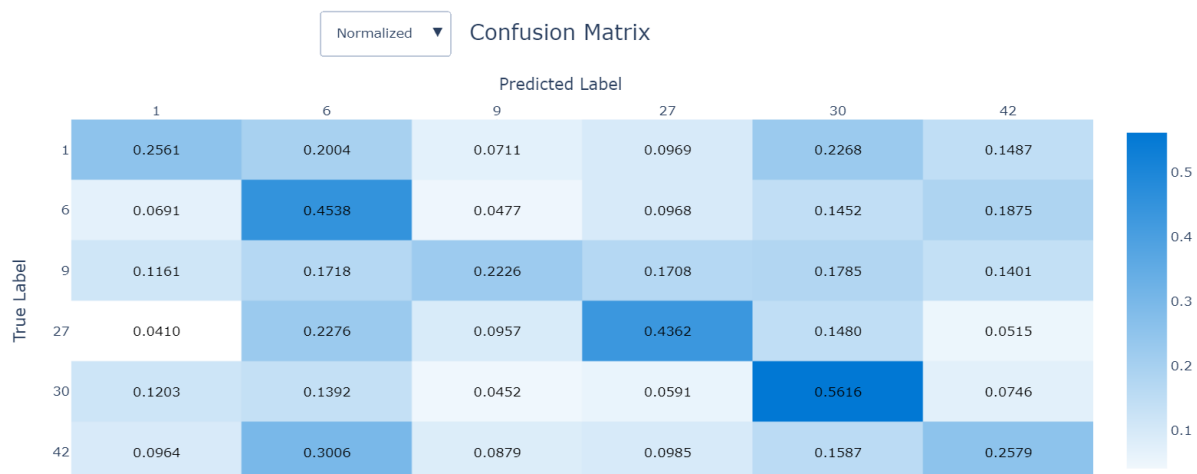
Tarkastellaanpa luokkien 6, 43 ja 44 onnettomuusprofiileita, verrokkina luokka 1 (musta käyrä). Kuten havaita saattaa, 43 ja 44 eivät juuri eroa toisistaan, jonkin verran eroa luokkaan 6, mutta ero luokkaan 1 on jo selvempi.



Kokeillaan mitä onnettomuusluokkien 43 ja 44, eli 2,3,4,5, yhdistäminen vaikuttaa, annetaan luokalle arvo 42 ja tietoaaineiston tekstitiedostolle nimeksi "datasetUusiOnLk2.csv".

Päivitetään työtilan Datasets osiossa jo oleva "datasetUusiOnLk" versioksi 2, klikkaa tietoaaineiston nimeä – New version – From local files. Käynnistetään uusi Automated ML ajo päivitetyllä tietoaaineistolla. Onnettomuusluokkajakauma näyttää tältä, nyt saatu kunnon epätasapainoa jakaumaan ja jännityksellä odotamme miten Automatic featurization hoitaa tilanteen.

Kulj_matk	(Multiple Items)		
Row Labels		Count of UusiOnLk	Count of UusiOnLk %
1	yksittäisonnettomuus	71025	17,31 %
6	peräänajo-onnettomuus	64999	15,84 %
9	jalanalkijajonnettomuus	11163	2,72 %
27	mopedi- tai polkupyöräonnettomuus	20212	4,92 %
7		21567	5,26 %
30	hirvi-, peura- tai muu eläinonnettomuus	21721	5,29 %
10		40901	9,97 %
11		2904	0,71 %
12			
42	kääntymis-, ohitus-, risteämis- tai kohtaamisenonnettomuus	38270	9,33 %
2		33314	8,12 %
3		61460	14,98 %
4		22861	5,57 %
5			
Grand Total		410397	100,00 %



Kuten huomaamme, yhä isoin osa luokan 42 ennusteista valuu luokalle 6, vaikkakin tilanne on hiukan tasahtunut. Myös luokan 9 ennustettavuus on saanut muutaman prosenttiyksikön parannusta aikaisempaan, joten Automatic featurization hoiti jälleen tasapainotuksen mallikkaasti.

Ennustemallin julkaisu

Seuraavaksi julkaisemme ennustemallin, vaikka ennusteen iterointia voisi toki vielä jatkaa esim. kokeilemalla jotain uutta muuttujaa raakadatasta, tällä erää tyydymme tähän. Valitaan "datasetUusiOnLk" version 2 ennusteaajoista paras malli, avataan ajon yksityiskohtainen näkymä Model Details – klikataan Deploy Model – annetaan nimi ja kuvaus, valitaan Compute: ACI (Azure Container Instance, toinen vaihtoehto Azure Kubernetes Service jos semmoinen konfiguroituna ja käytettävissä), Enable authentication valittuna - Deploy. Model Details näkymän kohdasta Deploy status voi seurata milloin malli on julkaistu. Kun valmis, voit ko. kentässä klikata mallin nimeä, avaa Endpoint osiossa kyseisen palvelinklusterin tiedot, vaihda Consume näkymään ja ota talteen Rest endpoint ja Primary key arvot seuraavaa vaihetta varten.

Deploy a model

Name *
tieliikenneonnettomuudet-uusilk2

Description
Tietoineistona UusiOnLk versio2

Compute type *
ACI

Models: AutoMLaee8574b948

Enable authentication
☒

This model supports [no-code deployment](#). You may **optionally** override the default environment and driver file.

☐ Use custom deployment assets

> Advanced

Preview

Microsoft Azure Machine Learning

New

Home

Author

Notebooks

Automated ML

Designer

Assets

Datasets

Experiments

Pipelines

Models

Endpoints

ML-Workspace2 > RealtimeEndpoints > tieliikenneonnettomuudet-uusilk2

tieliikenneonnettomuudet-uusilk2

Details

Consume

Basic consumption info

REST endpoint
http://[redacted]westeurope.azurecontainer.io/score

☒ Using key

☐ Using token

Primary key
[redacted] [Regenerate](#)

Secondary key
[redacted] [Regenerate](#)

Voit kutsua mallia REST API:n kautta käyttäen esim. Pythonia, joten luodaan Jupyter Notebook jolla kokeillaan mallin toimivuutta. Tähän projektiin liittyvä GitHub repo, jossa valmis Jupyter Notebook runko, löytyy täältä: <https://github.com/MarkoPeltojoki/TrafficAccidentsFI>

Hyvin malli näyttäisi toimivan.

```
# Create the strings from variables and output data with translations.
# datestring = datetime.today().strftime("%H:%M on %A, %B")
datestringFI = weathertranslateFI + ', kello ' + time + ' ja kuukausi ' + monthtranslateFI
datestringEN = weathertranslateEN + ' at ' + time + ' on ' + monthtranslateEN
accidentclassEN = translate['Classes'][output_data + 'c-en']
accidentclassFI = translate['Classes'][output_data + 'c-fi']
agestring = agerangetranslate

print('Accident class as output data: ' + str(output_data))
print('Prediction:')
print('Because you are in age range ' + agestring + ' and it is ' + datestringEN + ', the most probable traffic accident is in class ' + accidentclassEN + '.')
print('Ennuste:')
print('Koska olet ikäluokassa ' + agestring + ' ja on ' + datestringFI + ', todennäköisin liikenneonnettomuus on luokassa ' + accidentclassFI + '.')

Age range, weather, month and hour as Json input data: {"data": [[2, 5, 1, 13]]}
Accident class as output data: 6
Prediction:
Because you are in age range 18 - 27 and it is snowy at 13:46 on January, the most probable traffic accident is in class back collision accident.
Ennuste:
Koska olet ikäluokassa 18 - 27 ja on lumisade, kello 13:46 ja kuukausi Tammikuu, todennäköisin liikenneonnettomuus on luokassa peräänajo-onnettomuus.
```

```
Age range, weather, month and hour as Json input data: {"data": [[7, 2, 1, 13]]}
Accident class as output data: 9
Prediction:
Because you are in age range 85 - and it is cloudy at 13:55 on January, the most probable traffic accident is in class pedestrian accident.
Ennuste:
Koska olet ikäluokassa 85 - ja on pilvipouta, kello 13:55 ja kuukausi Tammikuu, todennäköisin liikenneonnettomuus on luokassa jalankulkijaonnettomuus.
```

Kun et enää tarvitse mallia tai sitä pyörittävää palvelinklusteria, poista se Endpoints osion kautta. Jos haluat säilyttää palvelinklusterin, mutta et käytä sitä pidempään aikaan, sammuta se Azure portaalin kautta, löytyy saman Resource Group:in alta minne loit Azure Machine Learning työtilan.

Yhteenveto

Näin olemme käsitelleet tässä artikkelisarjassa yhden ennustemallin tekemisen aina dataan tutustumisesta, esikäsitteystä, analysoinnista, jatkokäsittelystä, muuttujien etsinnästä ja valinnasta, luokkien hienosäätämisestä ja ennustemallien iteroinneista aina julkaisuun asti. Toivottavasti oli viihdyttävää luettavaa ja opit jotain uutta. Kiitän mielenkiinnostasi.

Tämän LinkedIn artikkelisarjan osat:

Osa1: <https://www.linkedin.com/pulse/ikä-tappaa-tieliikenteessä-osa-1-raakadata-marko-peltojoki>

Osa2: <https://www.linkedin.com/pulse/ikä-tappaa-tieliikenteessä-osa-2-datan-käsittely-marko-peltojoki>

Osa3: <https://www.linkedin.com/pulse/ikä-tappaa-tieliikenteessä-osa-3-ennustemalli-marko-peltojoki>