

Henkilöasiakkaiden laskumaksamisen mallintaminen ja ennustaminen aikasarja- analyttisin keinoin

Miikka Kirsilä

Prof. Pauliina Ilmonen

TkT Pekka Teppola

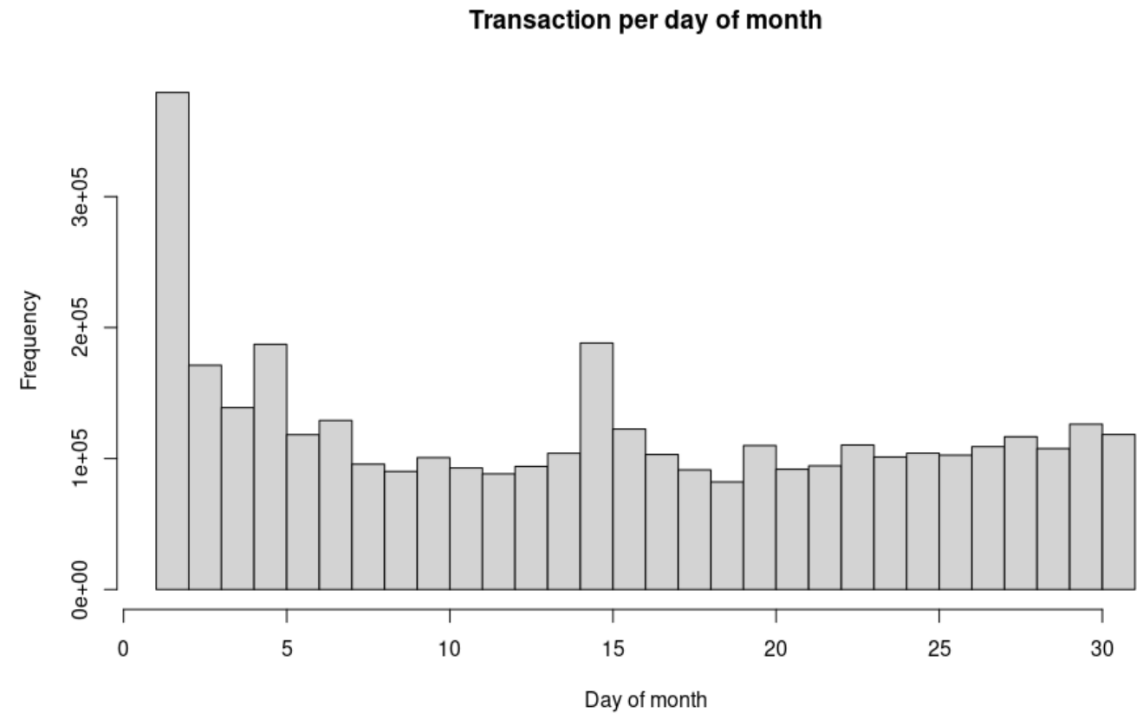
Tausta

- Tavoitteena ennustaa henkilöasiakkaiden kuukausittaisten laskujen summaa (euroina)
- Tällä pankki pyrkii:
 - Tarjoamaan parempia palveluita asiakkaille
 - Vähentämään asiakaspalvelijoiden manuaalisen työn määrää

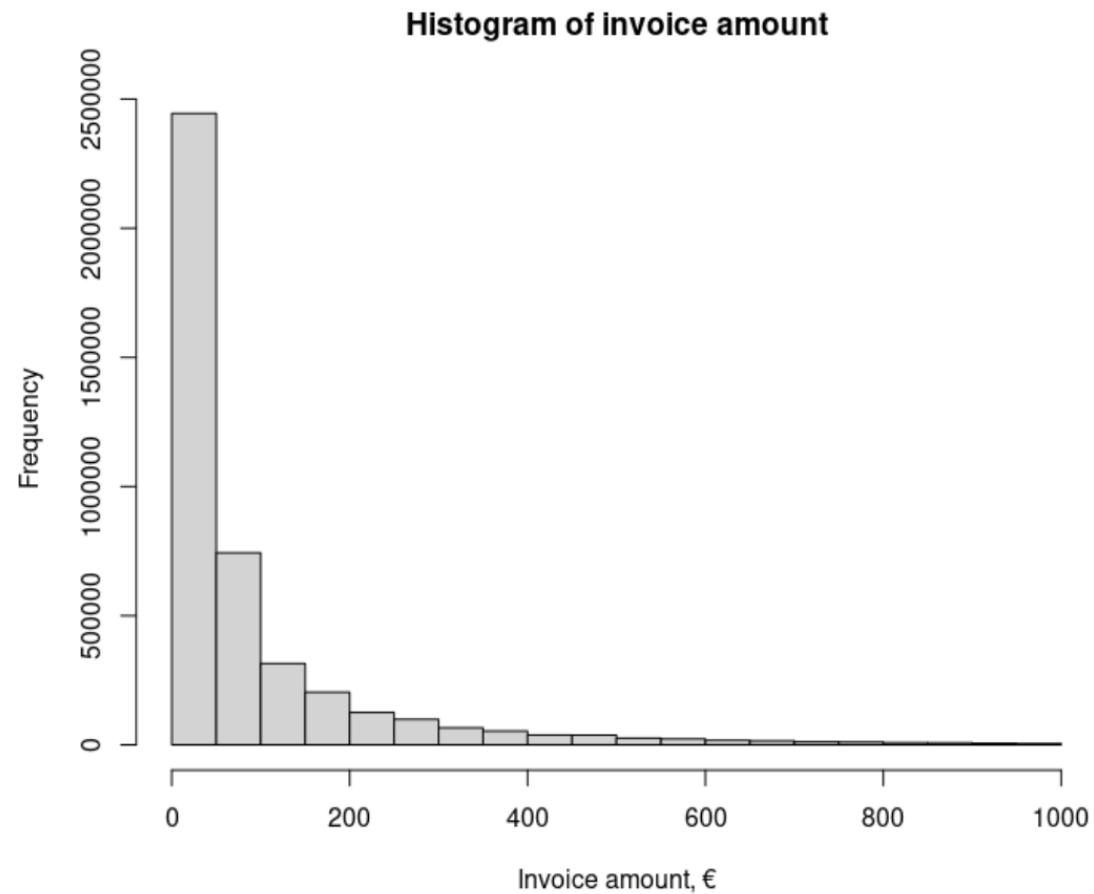
Valmis data

- Todellisten asiakkaiden transaktiodataa:
 - 3,5 vuoden ajalta: Tammikuu 2018 – Kesäkuu 2021
 - 4 327 994 transaktiota
 - 21 215 asiakasta
- Painotettu satunnaisotos todellisesta datasta

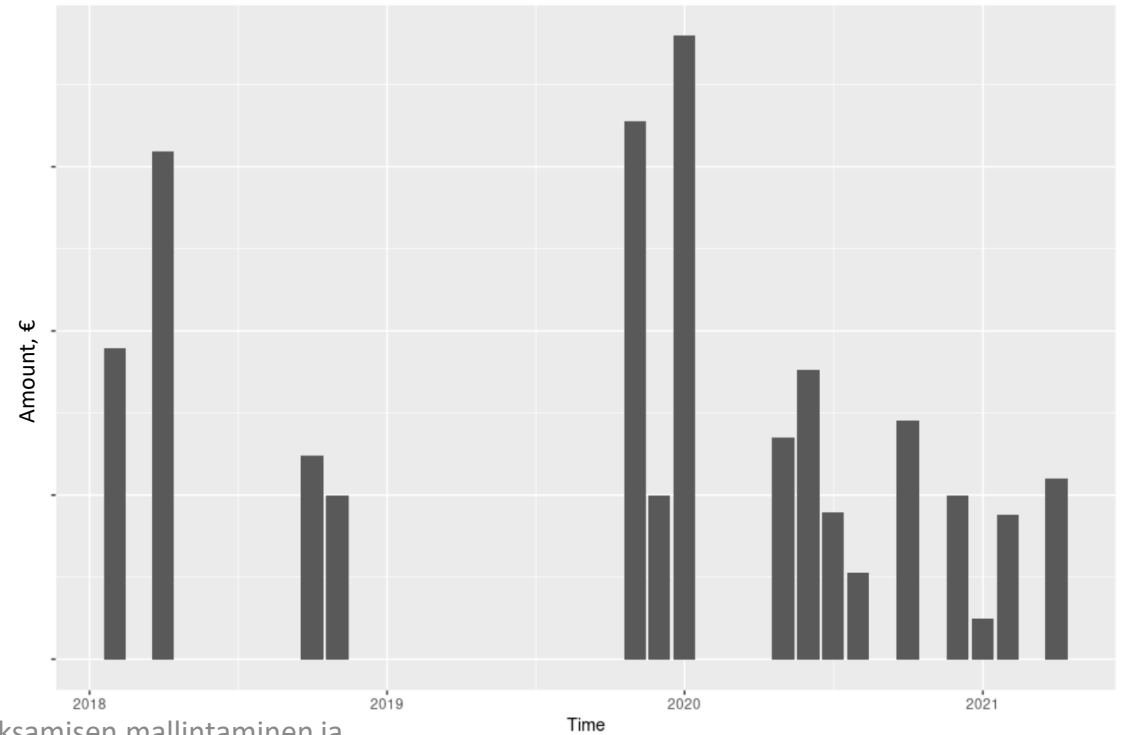
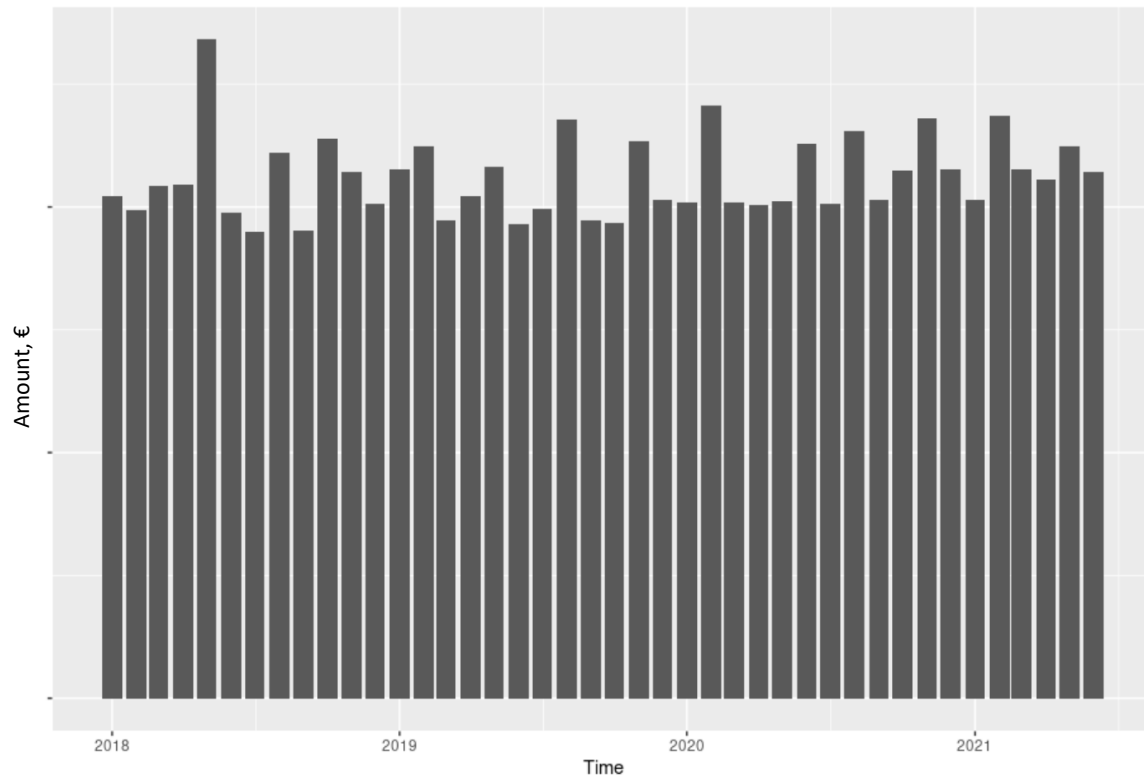
Transaktioiden lukumäärä



Transaktioiden arvo euroissa



Esimerkkiasiakkaita



Miten asiakkaiden
kuukausittaisten laskujen
summaa (€) voitaisiin ennustaa?

Kolme vaihtoehtoa

```
graph TD; A[Kolme vaihtoehtoa] --> B[ARIMA-malli]; A --> C[Facebook Prophet]; A --> D[Naivi lähestymistapa];
```

ARIMA-malli

Facebook Prophet

Naivi lähestymistapa

ARIMA-malli

Autoregressive:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \epsilon_t$$

Moving average:

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Integrated:

$$w_t = y_t - y_{t-1}$$

Facebook Prophet

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t)$$

Pitkän aikavälin trendi

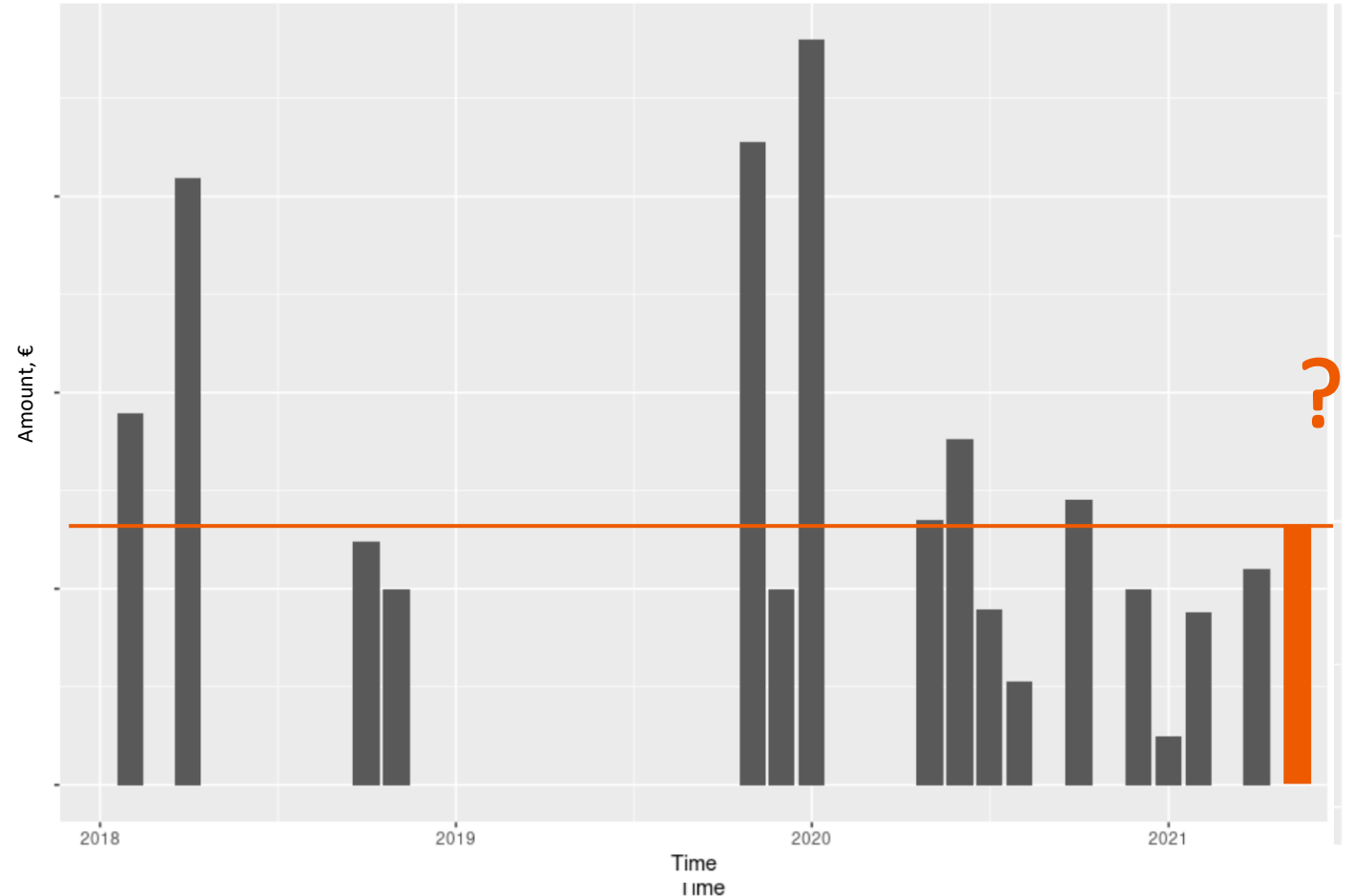
Kausittaisuus

Loma-ajat

Naiivi lähestymistapa

1. Laske keskiarvo muutamalta edeltävältä kuulta
2. Ennusta että seuraavassa kuussa menee keskiarvon verran rahaa

Nopea ja joskus tehokas malli, mutta ei aina paras.



Kysymyksenasettelu

1. Mikä edellämainituista malleista saavuttaa parhaat tulokset?
2. Saavuttavatko nämä mallit tyydyttäviä lopputuloksia?

Työkalut



Tulokset (1/5)

Table 1: Median and Mean of the MAPE-values of the ARIMA, Prophet and naive models with a test group of 50 persons on the second round of testing. Numbers rounded to five decimals.

	6 months	12 months	24 months	36 months
ARIMA median	0,82105	0,82054	0,80379	0,75718
Prophet median	NA	1,1526	0,7273	0,6135
naive median	0,9888	0,8939	0,4339	0,3525
ARIMA mean	1,48626	1,60351	1,27212	0,95904
Prophet mean	NA	2,19	2,029	1,5028
naive mean	2,0704	1,5212	0,9621	0,4995

Table 2: Summary of the RMSE-values of the ARIMA, Prophet and naive models with a test group of 50 persons on the second round of testing. Numbers rounded to three decimals.

	6 months	12 months	24 months	36 months
ARIMA median	655,107	659,29	688,536	453,787
Prophet median	NA	399,7	496,9	448,5
naive median	492,2	449,4	328,4	249,5
ARIMA mean	1200,349	1187,649	1161,505	930,699
Prophet mean	NA	984,5	1085,3	711,3
naive mean	1123,5	1019,7	691,3	403,9

Tulokset (2/5)

Parhaat ennusteet

Mitä jos valitaan jokaiselle käyttäjälle sopivin malli erikseen?

Tilastoidaan sopivin malli jokaiselle asiakkaalle



Valitaan sopivin historian pituus

Table 3: Summary of the MAPEs when the best model is chosen per individual customer.

Training data	min	median	mean	variance	max
6 months	0,04977	0,82105	1,48626	3,694194	11,40406
12 months	0	0,6381	0,9959	1,114902	4,9169
24 months	0	0,6336	1,0285	1,375415	6,2517
36 months	0	0,5243	0,7761	0,6219011	3,3688



Table 4: Summary of the MAPEs when the best model per individual customer is combined with optimal history length per customer.

	min	median	mean	variance	max
Combined best months	0	0,3486	0,4718	0,1065467	1,6043

Tulokset (3/5)

Parhaat ennusteet

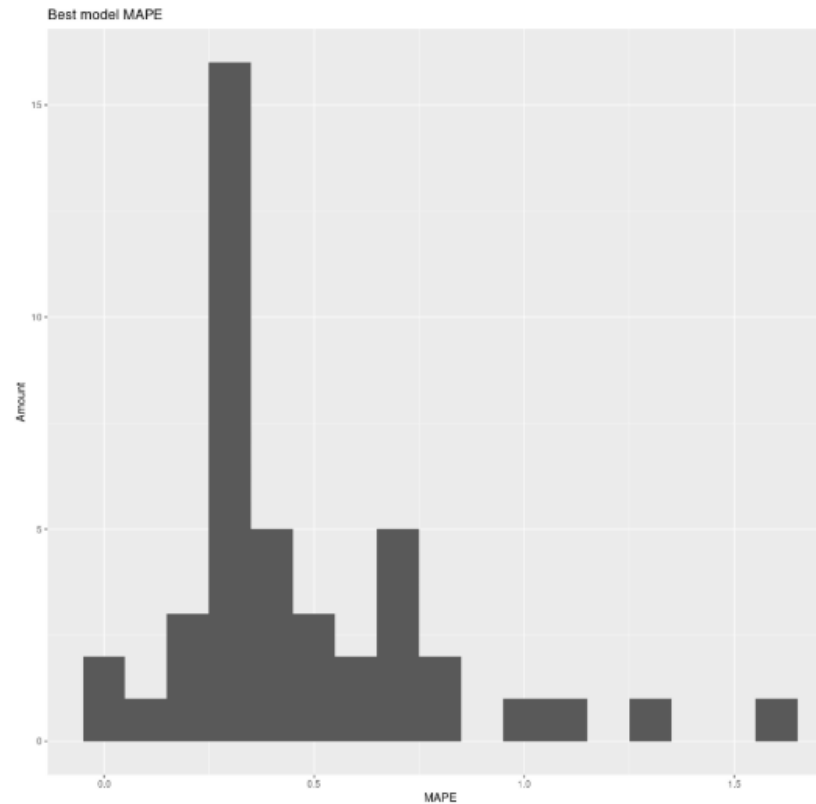


Figure 10: Histogram of the MAPE-values when the best model is chosen from a set of different training data lengths and the three models.

Tulokset (4/5)

Huonoimmat ennusteet

Tilastoidaan huonoin malli jokaiselle asiakkaalle



Valitaan huonoin historian pituus

Table 5: Summary of the MAPEs when the worst model is chosen per individual customer.

Training data	min	median	mean	variance	max
6 months	0,04977	0,82105	1,48626	3,694194	11,40406
12 months	0,04977	1,15265	2,79763	11,91872	17,76007
24 months	0,04977	0,87603	2,27264	6,950674	11,45
36 months	0,04977	0,81704	1,68574	4,218728	9,86518



Table 6: Summary of the MAPEs when the worst model per individual customer is combined with optimal history length per customer.

	min	median	mean	variance	max
Combined best months	0,04977	0,73311	1,17391	3,093816	11,40406

Tulokset (5/5)

Mallikohtaiset tulokset

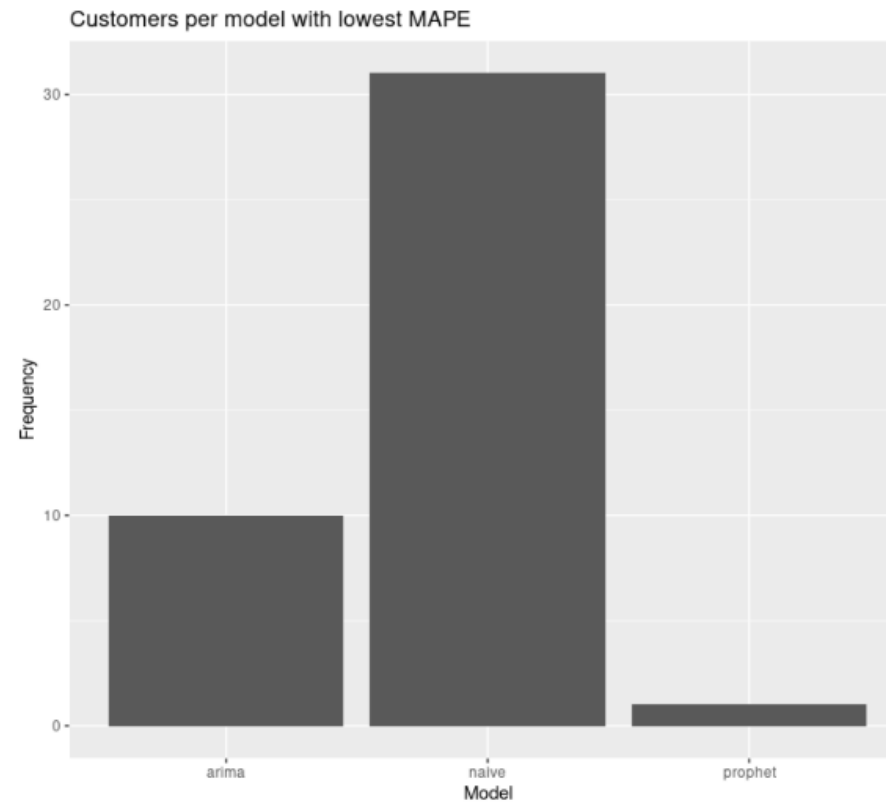


Figure 11: The amount of best predictions by model.

Tulevaisuus

ARIMA
+
Prophet
+
Naive



Lomakaudet sisältävä Prophet

Vaihtoehto `auto.arima()`:lle

Artificial Neural Networks

K-means clustering

Vastaanottajapohjainen ryhmittely

Kiitos