



Aalto-yliopisto  
Perustieteiden  
korkeakoulu

# Projection Predictive Method for Bayesian Model Selection in Retail Time Series Forecasting (Aihe-esitelmä)

*Elias Ylä-Jarkko*

*27.8.2021*

Ohjaaja: FM *Ella Tamir*

Valvoja: *Prof. Fabricio Oliveira*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

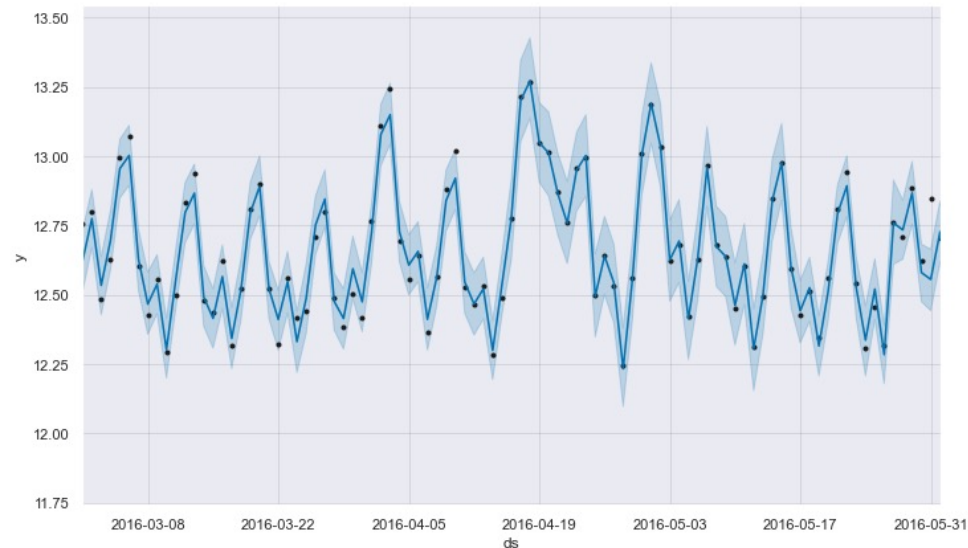
# Motivaatio

- Myynnin selittäminen tarjoaa taloudellista hyötyä
  - Johtuuko päivän myyntimäärä markkinoinnista, säästä, perjantaista, tms.
- Tarkempi malli kasvattaa hyötyä ja vähentää riskiä
- Muuttujien määrän vähentäminen voi lisätä tarkkuutta
  - Parantamalla ennustuskykyä
  - Lyhentämällä laskenta-aikaa
  - Helpottamalla päätöksentekoa

(Chan D. & Perry M., 2017), (Piironen, J. 2019)

# Menetelmä

- Ennustetaan equadorilaisen kauppaketjun myyntiä
- Muuttujina kausivaihtelu, öljyn, maissin ja perunoiden hinta, lämpötila, sademäärä, alennukset
- Referenssimalli: kaikki muuttujat mukana
- Sovitetaan yksinkertaisempia malleja referenssimallin ennusteeseen



# Menetelmä yksityiskohtaisemmin

- Valitsee relevantimpia muuttujia
- Stabiilimpi

(Piironen & Vehtari, 2017)

- Formaalisti: jokaista muuttujamäärää kohden valitaan se alimalli, joka eroaa täydestä mallista vähiten.

$$\begin{aligned}\theta_{\perp} &= \arg \min_{\theta \in \Theta} \text{KL}(p(\tilde{y} | \theta_*) \| p(\tilde{y} | \theta)) \\ &= \arg \max_{\theta \in \Theta} \mathbb{E}_{\tilde{y} | \theta_*} (\log p(\tilde{y} | \theta)).\end{aligned}$$

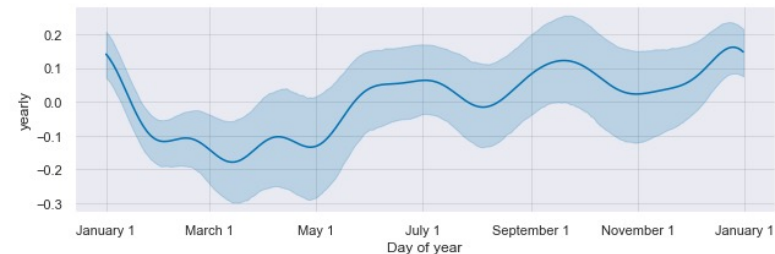
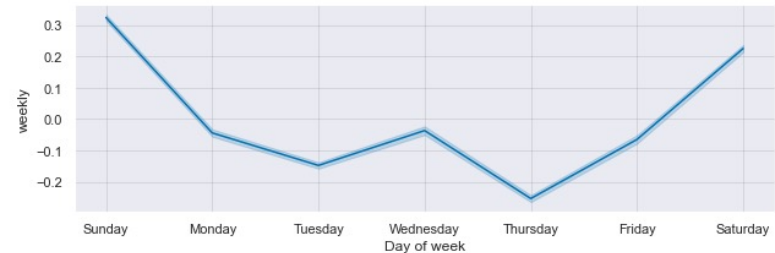
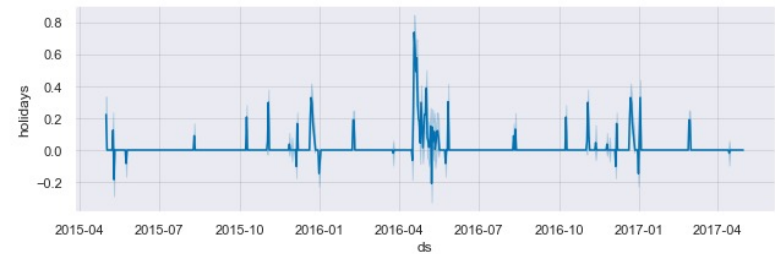
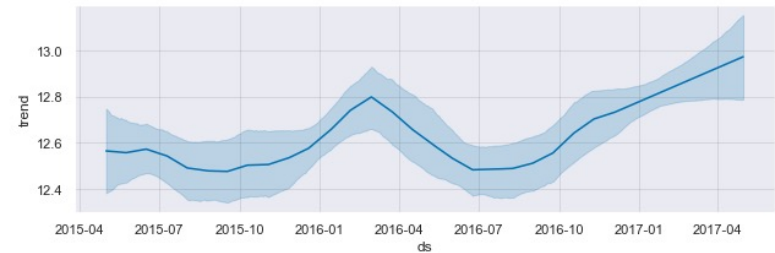
- KL-divergenssin minimoiminen on ekvivalenttia täyden mallin ennusteisiin MLE-sovittamisen kanssa.

# Työkalut

- Python:lle ja R:lle toteutettu aikasarjamallintamisen kirjasto
- Taustalla Stan
- Sovittaa automaattisesti jaksollisuuksia, paloittaisen trendin ja yksittäisiä lomapäiviä

$$y \sim N(\text{trendi} + X * \beta, \sigma)$$

$\beta$  = ”muuttujat ja fourier–kertoimet”



# Tavoitteet

1. Tutustua siihen miten Projection Predictive Method soveltuu vähittäiskaupan aikasarjamallintamiseen Prophet-kirjastolla
2. Selvittää millaiset muuttujat selittävät vähittäiskaupan kokonaismyyntiä parhaiten

(Lisäksi) Selvittää miten Prophet-malleja täytyy muokata jotta ne soveltuvat menetelmään.

# Aikataulu

- 6-7/2021 Aiheen valinta, datan etsiminen ja puhdistaminen, työkaluihin tutustuminen
- 8/2021 Projektimenetelmän ohjelmoiminen ja tulosten kerääminen
- 9/2021 Yhteenveto, kirjallisuus

# Lähteet

- Chan, D., & Perry, M. (2017). Challenges and opportunities in media mix modeling.
- Piironen, J. (2019). Bayesian Predictive Inference and Feature Selection for High-Dimensional Data, *Aalto University publication series DOCTORAL DISSERTATIONS*, 84/2019
- Piironen, J., & Vehtari, A. (2017). Comparison of Bayesian predictive methods for model selection. *STATISTICS AND COMPUTING*, 27(3), 711-735. <https://doi.org/10.1007/s11222-016-9649-y>
- Piironen, J., Paasiniemi, M., Vehtari, A., (2020) Projective inference in high-dimensional problems: Prediction and feature selection, *Electronic Journal of Statistics*, 14/2020 2155–2197