

1|

Throughput Analysis

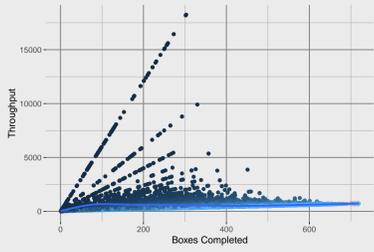


Figure: Throughput vs Boxes Completed (colour is run time)

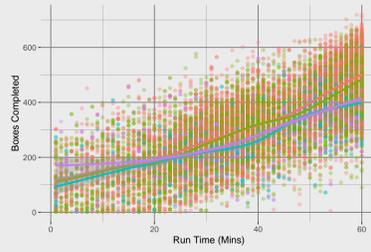


Figure: Boxes Completed vs Run Time for all production lines

- Throughput = Produced Units per unit Run Time
- Understanding throughput is crucial for design and operation of supply chains
- Currently operating 2 factories, 4 production lines. With 2 more factories opening soon

Aim: Understand what influences throughput

Aim: Does the solution translate?

Aim: Can changepoint analysis extract context?

2|

Challenges

Challenges:

- Modelling non-linear problem in simple, explainable terms
- Assessing translatability between production lines
- Accounting for missing, erroneous and outlying values
- Building Random Forest's and optimising hyperparameters

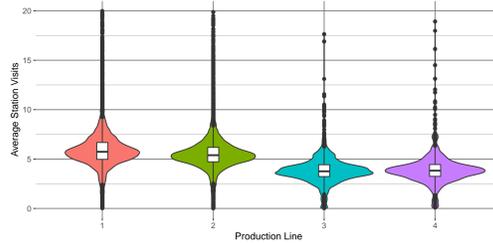


Figure: Average Station Visits for each production line trimmed at $x > 20$ (Extremely Heavy Tailed dataset)

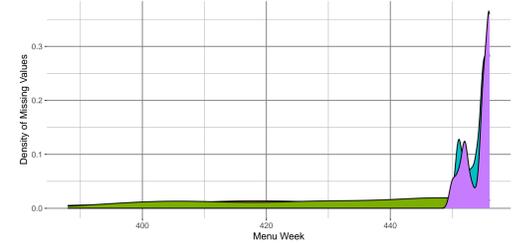


Figure: Missing Values by Production Line. Lines 3 and 4 see significantly more missing values

3|

Random Forests

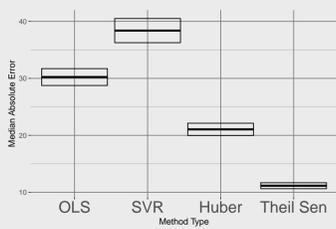


Figure: MedAE Kfold $k = 15$ Validation Mean and SD

- Building pipelines to simplify the process of creating, training and testing models
- Analysing various robust approaches to deal with outliers [(P. Huber, 2009)]
- Tuning hyperparameters using multiple methods to extract maximum information
- Convert the best random forest to a kernel to simplify the complex non-linear relationships.

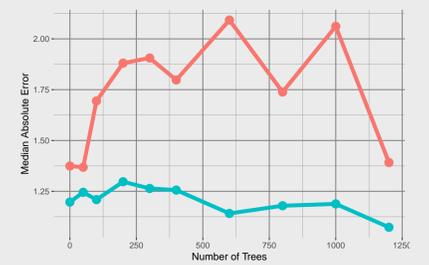


Figure: MedAE for KeRF (red) vs RF (blue) for n estimators

4|

Changepoint Analysis



Figure: Median Hourly Boxes Completed per Month. $n = 3$ changepoints, demonstrating the clear 3 phase pattern with a peak in the middle dropping with the opening of lines 3 and 4

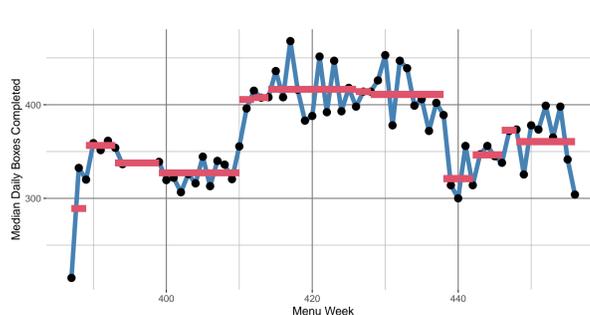


Figure: Median Hourly Boxes Completed per Menu Week. $n = 12$ changepoints, showing the 3 phase pattern

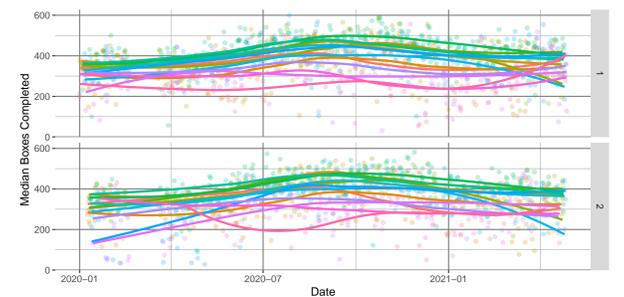


Figure: Median Hourly Boxes Completed per Shift Sequence. Shifts later in the menu week demonstrate significantly lower production than those earlier - but this changes over time with a change in scheduling algorithm

5|

Take-aways

Summary

1. Random Forests achieve 96.17% % Deviance, and Well-Optimised RF increases to 96.62%. With MedAE going from 1.179 \rightarrow 1.073
2. KeRF reduces the accuracy of the RF model (1.073 \rightarrow [1.2, 2.1]), but increases interpretability
3. Model is translatable and Changepoint Analysis provides increased contextual information for the regression models

Further work

- Assessing performance using totally unseen-data to evaluate forecasting potential
- Utilising un-summarised data to give more accurate insights
- Exploring the application of changepoint analysis with further contextual industry knowledge

Contact



LinkedIn: @kbmolloy
Github : @K-Molloy