Should feature selection be part of the Pipeline?

Feature selection in the pipeline

If deploying for the first time,

Plus

- We don’t have to hard code the predictive features

However,

- Need to deploy code to engineer **all features** in the dataset
- Error handling and unit testing for all the code to engineering features

Feature selection in the pipeline

What if we re-train our model frequently?

Advantages

- Can quickly retrain a model on the same input data
- No need to hard-code the new set of predictive features after each re-training

Disadvantages

- Lack of data versatility
- No additional data can be fed through the pipeline

Feature selection in the pipeline

In summary,

Suitable:

- Model build and refreshed on same data
- Model build and refreshed on smaller datasets

Not suitable,

- If model built using datasets with a high feature space
- If model constantly enriched with new data sources