**The Problem:**

Frequency in Process data is significantly higher than the Product Data.

**Question of Interest:**

Determination of the most important process parameters (or variables). Alternatively, the variables having highest impact on the Product quality.

Process Parameters

Observation (On time scale)

Product Parameters

Observation (On Product Features)

Process Data

Product Data

**Challenge:**

As our question of interest is determining the process parameters with highest impact on product quality. Therefore, we need to relate both data sets (Process and Product). Due to significantly different frequency of collection, it is not straightforward to merge these two data sets.



Process Parameters

Observation (On time scale)



Product Parameters

Observation (On Product Features)

**Solution:**

To overcome the problem of different collection frequency we purpose a method to summarize process data (data with high frequency) with respect to summary some functions (f{x}) such as mean, median, Mode etc.

Often process data is collected on time scale i.e features such as ARIMA components may also be very relevant to understand the distribution of the data.

Process Parameters

Observation (On time scale)



Product Parameters

Observation (On Product Features)

f{x}

Process Data (Summarized)

The small green box are the part of the process data (In our case layers).

**Solution:**

On x-axis of the below diagram we will have different functions (mean, median, mode, AR and MA) and on y-axis we will have different Product parameters (`Young's modulus (MPa)`, `Tensile strength (MPa)` and `Strain at break (% strain)` or more ).

Process Data (Summarized)

Product Feature 1

Process Data (Summarized)

Product Feature 2

Process Data (Summarized)

Product Feature n

Process Data (Summarized)

Product Feature 1

Process Data (Summarized)

Product Feature 2

Process Data (Summarized)

Product Feature n

Process Data (Summarized)

Product Feature 1

Process Data (Summarized)

Product Feature 2

Process Data (Summarized)

Product Feature n

The Idea is to run prediction algorithm such as XGBOOST for each element of the matrix and calculate the importance matrix (presented as yellow vector).

Process Data (Summarized)

Product Feature 1

Process Data (Summarized)

Product Feature 2

Process Data (Summarized)

Product Feature n

Process Data (Summarized)

Product Feature 1

Process Data (Summarized)

Product Feature 2

Process Data (Summarized)

Product Feature n

Process Data (Summarized)

Product Feature 1

Process Data (Summarized)

Product Feature 2

Process Data (Summarized)

Product Feature n

As results for the most important variable we purpose a criteria where all vectors with variable importance information (yellow vectors) will be ranked with respect to model fit measure (the blue vector).