DISTILLATION COLUMN BEHAVIOR MODELING

Anto Francis
DAT-5, General Assembly

CONTENT

BACKGROUND:

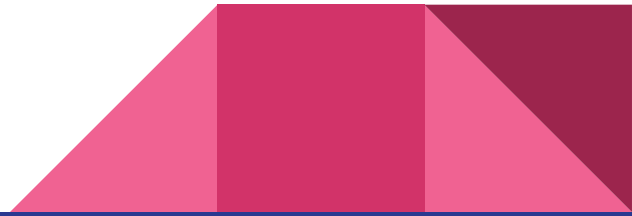
- What is Distillation Column
- Benefits of Good model

CURRENT INDUSTRY SITUATION & SHORTCOMINGS

PROJECT OBJECTIVE

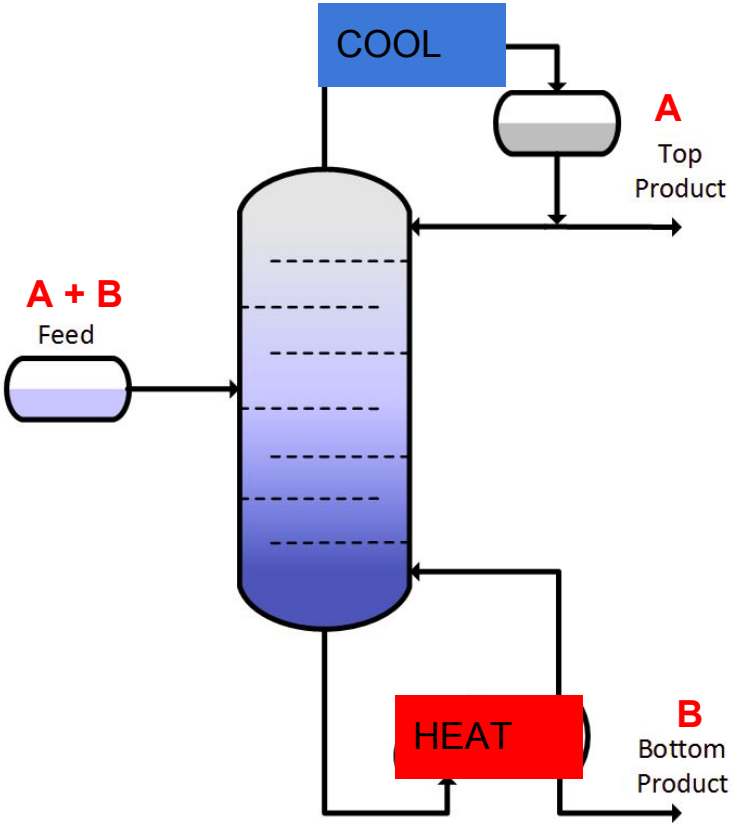
METHODOLOGY

RESULTS & INFERENCES



BACKGROUND

WHAT IS DISTILLATION COLUMN



Most ubiquitous equipment for separation of compounds in process industry

Separates compounds based on its difference in boiling points (tendency to vaporize)

A petrochemical process plant consumes
~ 200 - 800MW energy
~ 50 - 200Mn$/yr

BENEFITS OF GOOD MODEL

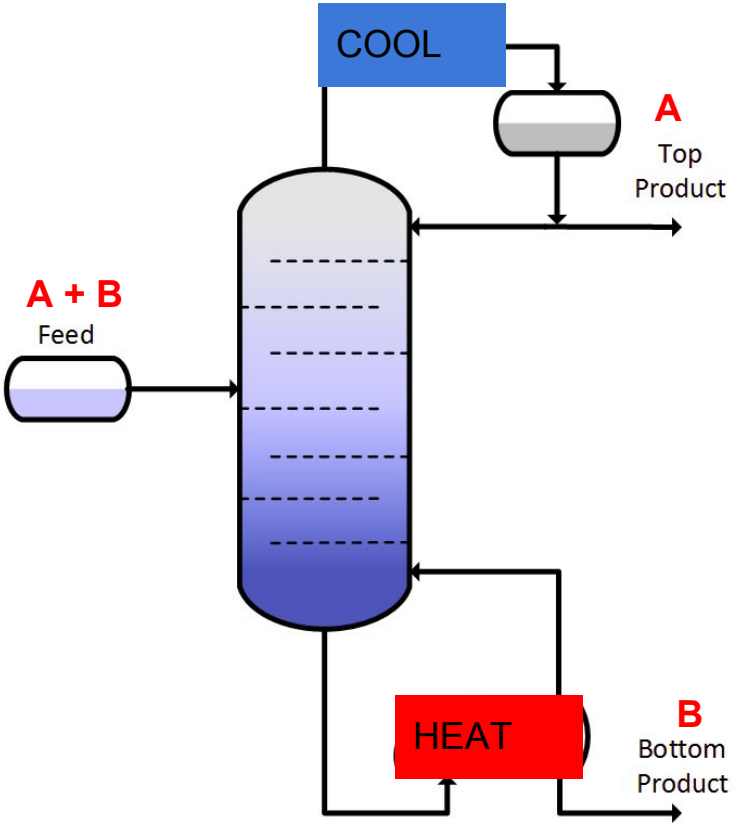
Column behavior models are stored in softwares to enable operating a column in its most optimal state

Even 2% energy efficiency improvement translates to 1 to 4Mn$/yr per process plant

Thus, significant economic benefit exists to operate the column at its optimal state

CURRENT INDUSTRY SITUATION

CURRENT DISTILLATION COLUMN METHOD



At present there are linear models which approximate Equipment behavior

Factors influencing product quality:

- Heat at bttm (Q)
- Cooling at top (R)
- Feed Rate (F)
- Feed Comp (FA)

Some influencing factors like climate, cannot be measured. Hence, not fed into model

Equipment control software uses model to maintain product quality & minimize energy usage

CURRENT METHOD - SHORTCOMINGS

MAJOR ISSUES:

- Equipment behavior changes over time & model becomes outdated
- There are unmeasured disturbances (e.g rain) which impact the equipment behavior and invalidates model behavior

AUXILLARY ISSUES:

- Equipment behavior is non-linear. Thus, linear models unable to extrapolate accurately
- Instrumentation drift and start lying
-

PROJECT OBJECTIVE & METHODOLOGY

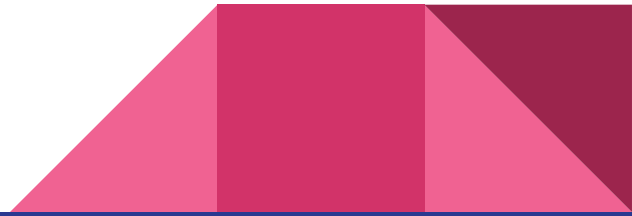
PROJECT OBJECTIVE

Compare 2 Time Series Model performance:

- Linear (Multi-variate AR)
- Neural Nets (Multi-variate LSTM)

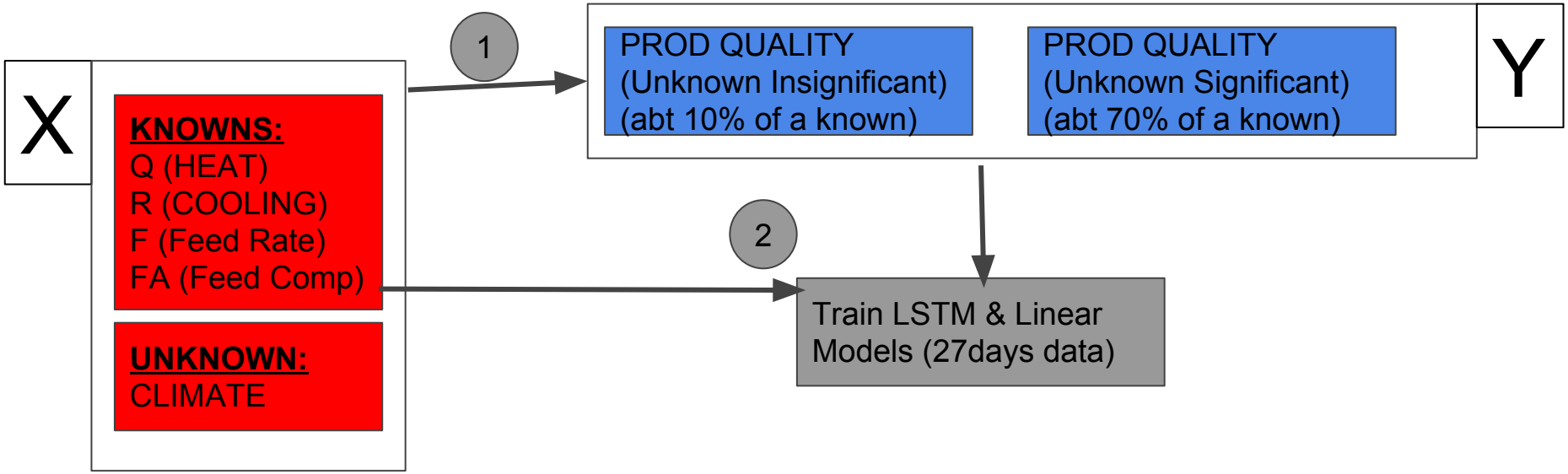
Compare Model performance in 2 different scenarios:

- Unknown variable has low impact
- Unknown variable has high impact

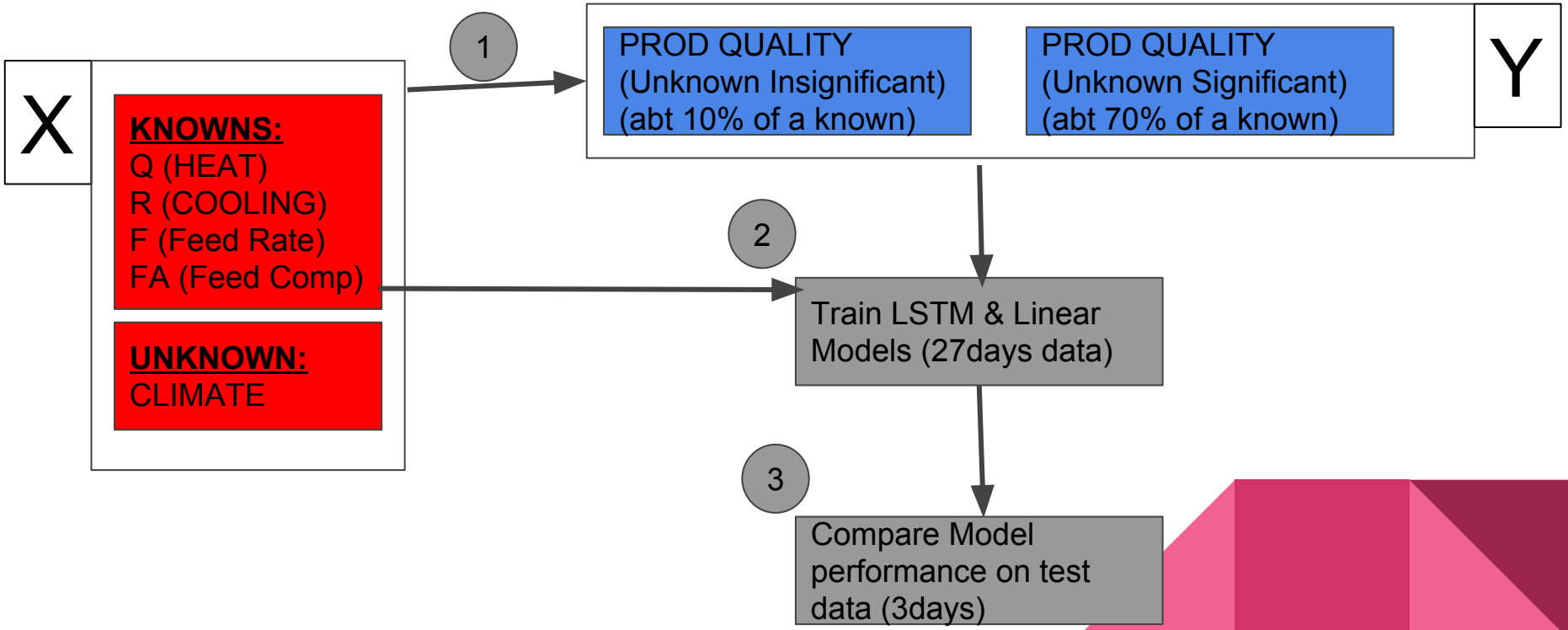


METHODOLOGY - Step1 (Generate Data)

METHODOLOGY - Step2 (Train Models)

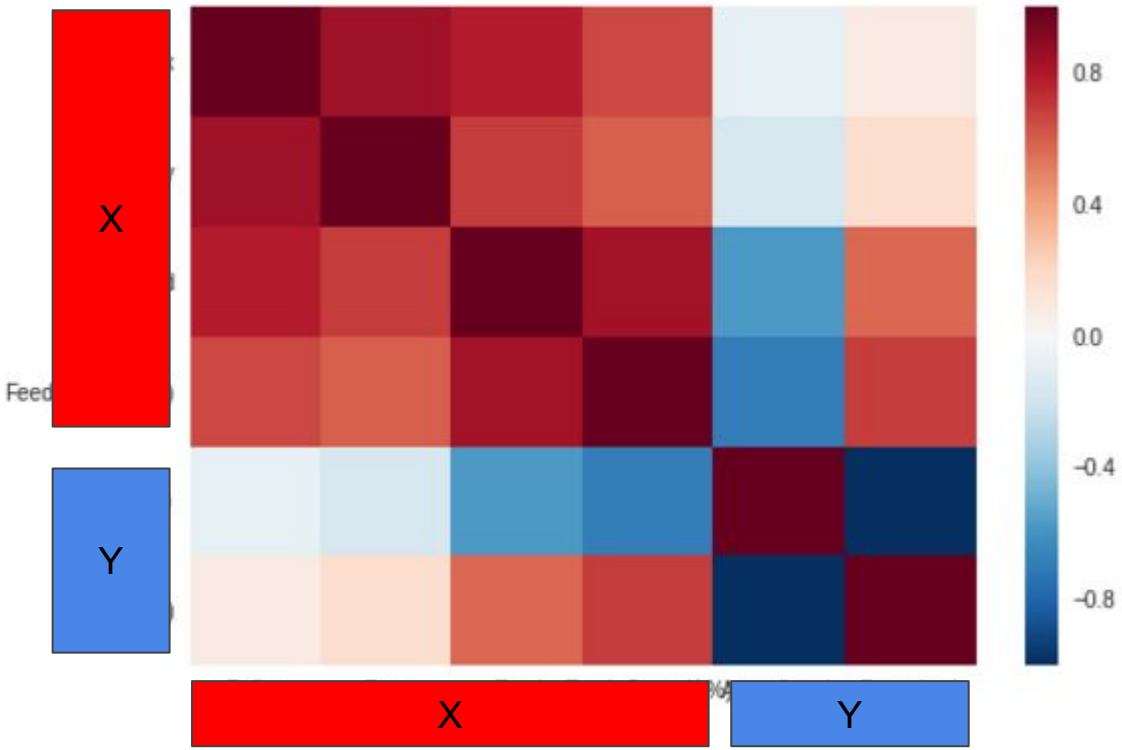


METHODOLOGY - Step3 (Test Model)



RESULTS & INFERENCES

Generated Data Characteristics



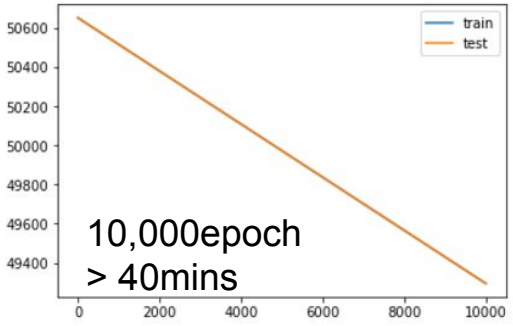
The 2 Quality streams are highly correlated

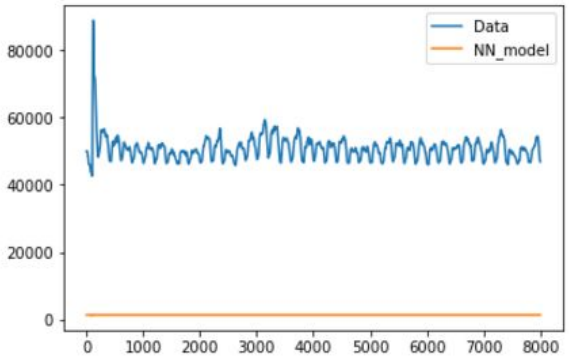
Inputs are highly correlated as expected from plant data

we try to maintain quality in a narrow band - correlation b/w input variables expected

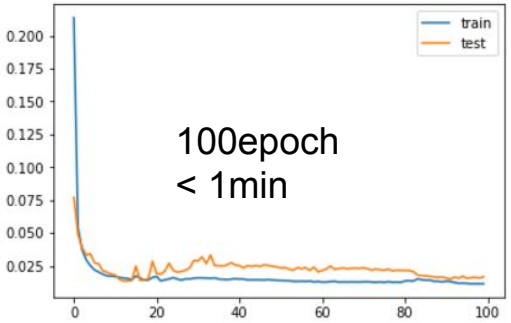
MODEL TRAINING

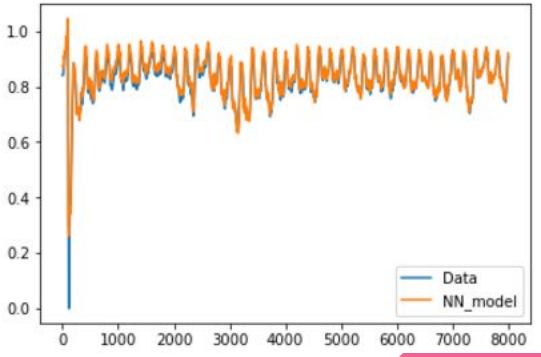
WithOut Scaling





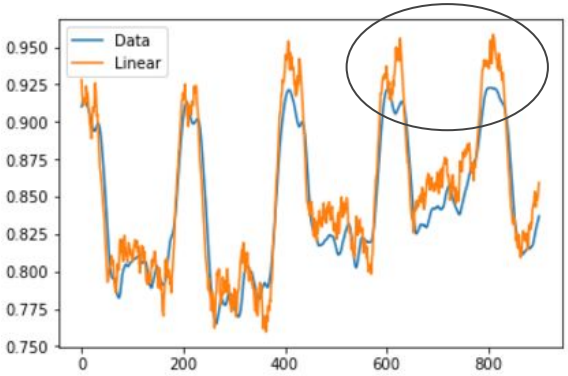
After Scaling





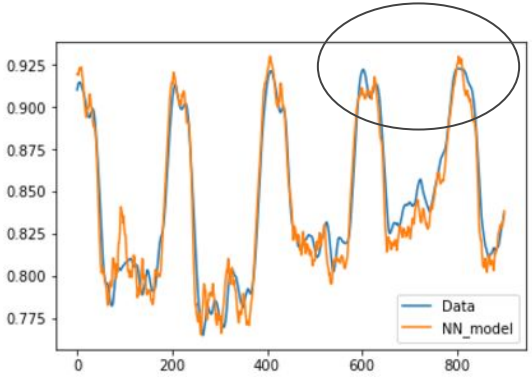
Linear vs NeuralNet - Unknown Insignificant

LINEAR



	MSE (1e-4)
Train Data	9.3
Test Data	3.0

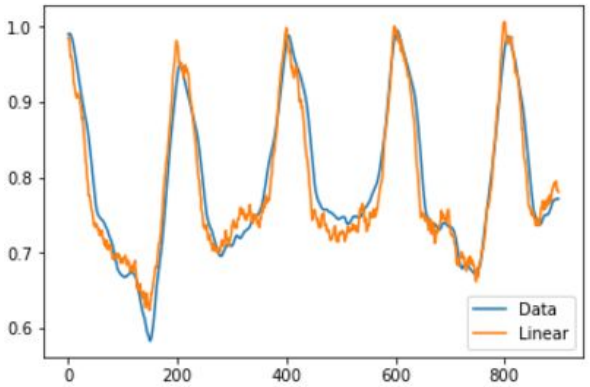
LSTM



	MSE (1e-4)
Train Data	7.4
Test Data	1.3

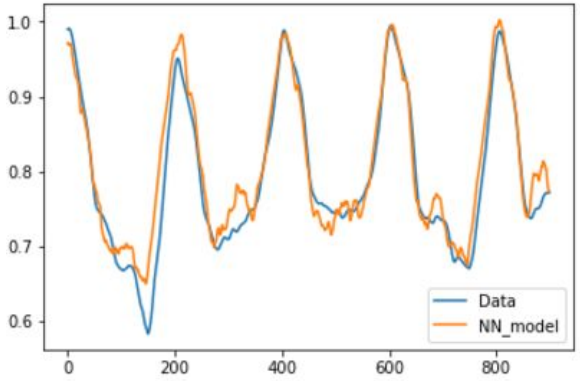
Linear vs NeuralNet - Unknown Significant

LINEAR



	MSE (1e-4)
Train Data	16.8
Test Data	7.5

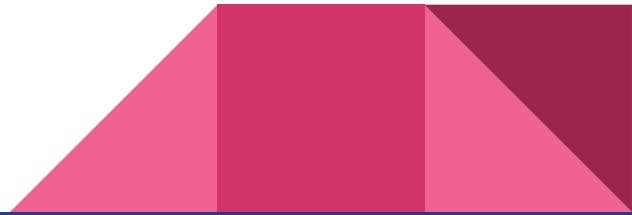
LSTM



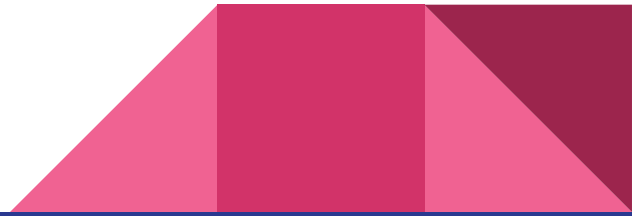
	MSE(1e-4)
Train Data	14.6
Test Data	8.9

INFERENCE & SUMMARY

- When Unknowns has significant influence - Linear Models are better
- When Unknowns don’t have significant influence - Neural Nets are better
- As Neural nets try to explain the effect of unknown variable using the knowns, it seems to over-fit and degrade the model
- Probably, a better self-learning model can be used with following approach:
  - Neural-Nets with periodic model learning
  - When indications of unknown variables being active - stop model learning process



THANK YOU



MEHODOLOGY

1. Generate Data Based on engineering Judgement
2. Compare metrics to ensure data representative of expectation
3. Input / OutPut Model
    a. X : Heat; Cooling; FeedRate; FeedComp; Unknown
    b. Y : Quality of 2 streams
4. Compare test data prediction accuracy b/w 2 models over varying degree of influence by the unknown variable