

Building Robust Processes through Application of Low Code ML & AI Models

Anurag Seksaria

Managing Partner, Anurag Seksaria Management Consulting, Bangalore, India.

E-mail : hello@anuragseksaria.org

Abstract

The authors will like to present an approach to build robust processes using Low Code Machine Learning and Artificial Intelligence Models specifically in continuous industry like Chemicals, Cement, Metals where major component of raw material is natural resource like Iron Ore, Bauxite, Coal, Lime where the quality parameters have high and uncontrollable variations.

The presentation is based on authors experience of working with some of the large Indian industry over last couple of years and his own research on the subject.

Like any other industry, the quality parameters like Purity, Yield, Moisture Content are function of Raw Material Quality (XRM), Process Parameters (XPP) and Intermittent Product Parameters (XIP)

$$Y \text{ (Output Parameter)} = f \{ X_{RM1}, X_{RM2}, \dots, X_{PP1}, X_{PP2}, \dots, X_{IP1}, X_{IP2}, \dots \} + \epsilon$$

Author has tried the popular UOSEMN model to create predictable models and deploy them in manufacturing processes to enhance Quality and Productivity with minimal human interventions.

Understand Business Requirement: While it appears obvious, its of utmost importance to understand the expectation of the business stakeholders.

Obtain Data: Understand and compile data for end-to-end process, starting from Raw Material, In-process Checks and Finals validation of the products.

Scrub: Although both quality of and quantum of the data available has improved many folds, a careful scrubbing of data is essential to getting predictable models. Any short cut during this stage may cost multiple iterations and delay in getting the desired results. Continuous process industry has additional constraints in terms of

- Time Lag between the processes
- Heterogeneity in the raw material, in process and final product
- Collinearity, that is relationships between the influencing factors.

Model: Build Predictable Machine Learning Models using methods like Logistics Regression & Decision Tree. No Low or No Code approach helps in engaging domain experts in building their own models which can be generalized but remain flexible.

iNterpret : The business insights are drawn and processes are optimized with close collaboration of management teams. Post the pilot runs, AI algorithms are created to deploy the model and accomplish in process adjustments.

Author experiences large scale deployments resulting in transformational improvement in Efficiency, Effectiveness and Stakeholder Experience.

Key Words : Low Code Machine Learning Models, Time Lag, Heterogeneity, Collinearity, Decision Tree, Logistics Regression-OSEMN Model

I. INTRODUCTION

Large number of process industries like Chemical, Pharma, Metals (both Ferrous & Non-Ferrous, etc. are working n optimization of the quality parameters like Purity, Yield, Moisture Content to improve the efficiency and effectiveness of the business.

These parameters depend on Raw Material Quality (XRM), Process Parameters (XPP) and Intermittent Product Parameters (XIP). While the scenario to any other industry, some of the challenges faced by this industry are unique.

- Time Lag between the processes
- Heterogeneity in the raw material, in process and final product
- Collinearity, that is relationships between the influencing factors.

To be specific two of the important parameters in some of the natural Raw Materials is number of Fines and Amount of Coarse Material which are directly related to each other and will always have high negative correlation within a range.

The challenging aspect here is that the process parameters are continuously fine-tuned in the control room to optimize the process given the changes in the ratio of course to fine inducing Auto Correlation in the system. Some time it happens through a well defines rule within the control room systems, while at many occasions it happens through interventions by process engineers.

II. UNDERSTAND BUSINESS REQUIREMENT:

The broad objective of the exercise was to Maximize Yield and build a generalized model which can be used to optimize process during day to day operations.

$$(\text{Output Parameter}) = f \{ X_{RM1}, X_{RM2}, \dots, X_{PP1}, X_{PP2}, \dots, X_{IP1}, X_{IP2}, \dots \} + \epsilon$$

Figure 1 : Distribution for Yield

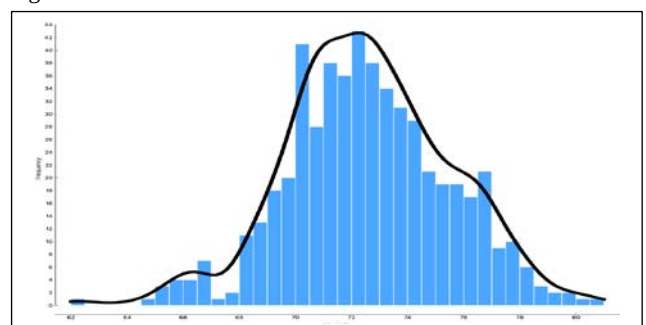
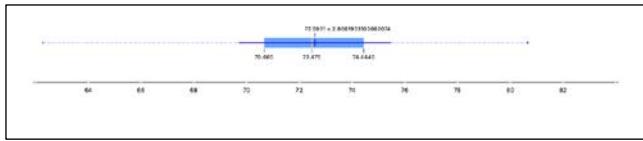


Figure 2 : Box Plot for Yield



The Distribution and Box Plot show that there is large variation in Yield over given period. There are also outliers on both lower and higher end of the observations.

Further analysis showed that the lower end outliers were due to certain abnormalities in the process like bathes post shutdown, machine breakdowns etc., hence the outliers on lower side were removed from the analysis.

Since the extreme values appears to be part of the multi modal distribution which is emerging due to efforts to improve yield were retained for the analysis.

II. OBTAIN DATA

The data was obtained for around 18 months where the yield was measured once every day.

Post this the type of variable and role for each of the parameter were decided.

The details of the features considered is given in Annexure 1.

III. SCRUB

To large extent the data was already clean due to earlier initiatives, Feature Construction was tried with Coarse to Fine Ratio as the Feature rather than considering them independently.

IV. EXPLORE

The feature statistics for the observed parameters in shown in Annexure 2. Further Box Plots and Scatter Plots were used to identify relationships between various parameters. Some of it is shown in Figure 3,4,5 and 6.

Figure 3 : Yield by Dosing Rate

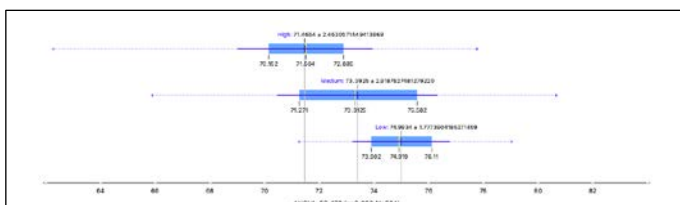


Figure 4 : Yield By A Feed

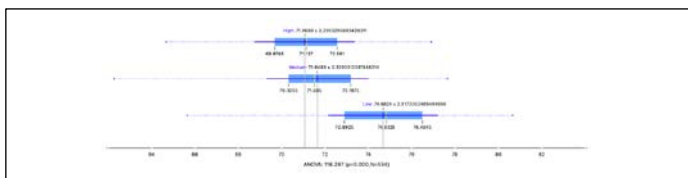


Figure 5: Scatter Plot Between Yield & Spray 1 with A Feed

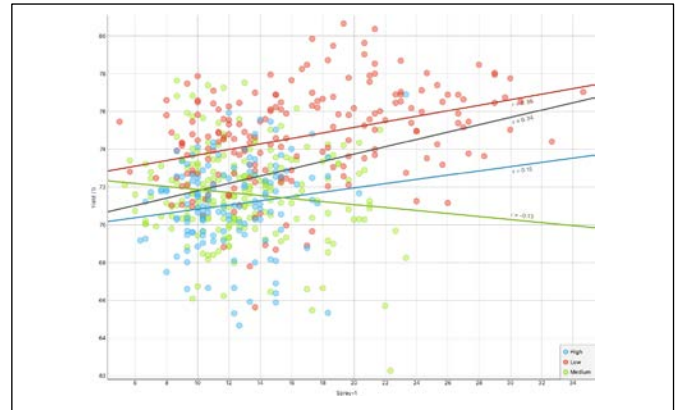
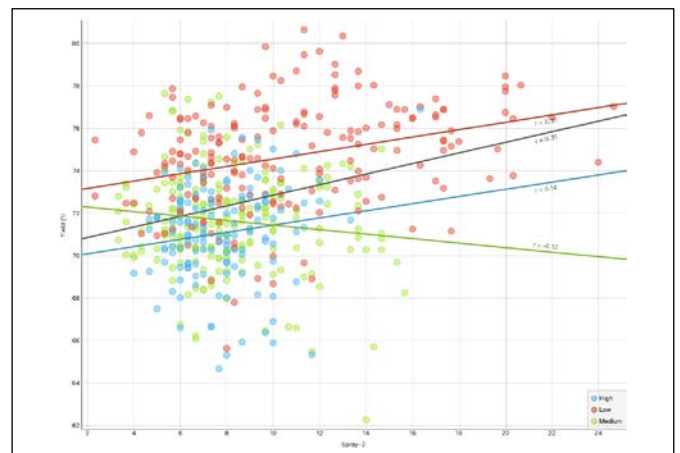


Figure 6: Scatter Plot Between Yield & Spray 1 with A Feed



While Figure 3 and 4 shows the difference in Yield with changes in Dosing Rate and A Feed, The scatter plots show in Figure 5 and 6 show very interesting finding.

The direction of change in Yield for Spray one changes with Dosing Rate. For Dosing Rate Medium there is a near ZERO impact on Yield with Change in Spray 1.

Whereas when Dosing Rate if High or Low there is a positive relationship between Yield and Spray 1 with an R of +0.3.

V. MODEL

Modeling was carried out using Decision Tree, Linear Regression and Random Forrest with Bootstrapping and 20 Cross Folds. While R-Sqr was the initial measure for model selection, the final test was assessing residual which is discussed in the later part of the paper.

Figure 7 : R-Sqr, MSE, RMSE, MAE for various models

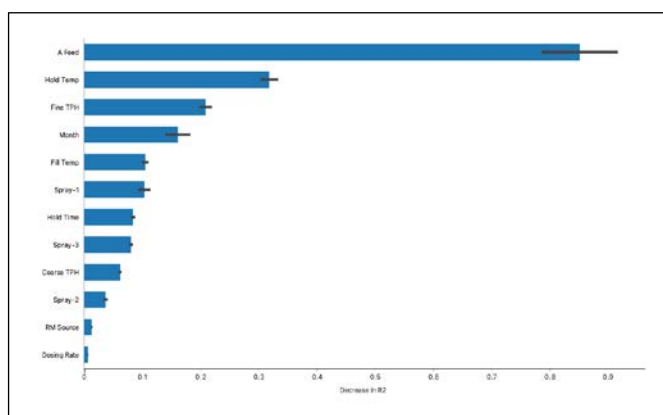
Model	MSE	RMSE	MAU	R-Sqr
Random Forest	2.3	1.3	1.0	73%

Tree	3.2	1.8	1.3	62%
Linear Regression	4.2	2.1	1.7	50%

VI. INTERPRATE

The next step was to find out feature importance with respect to impact of R-Sqr. These are given in

Figure 8.



The optimal combinations which maximizes yield to around 80% from the base value of 72% were identified using SHAPLY Plots Figure

Figure 9

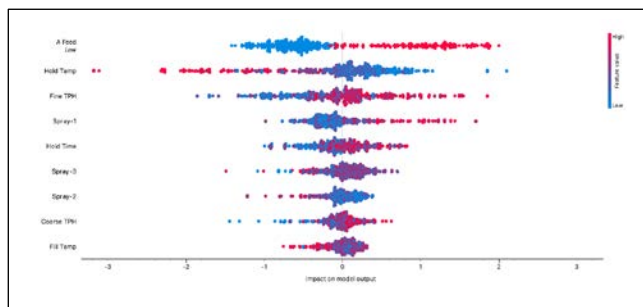
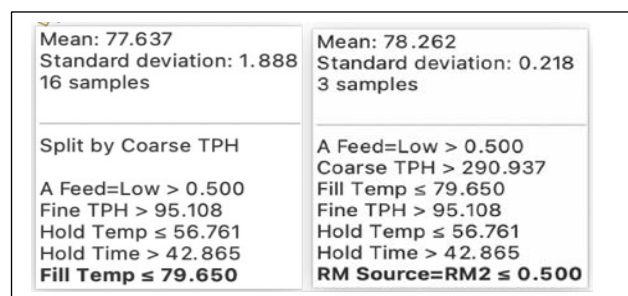


Figure 10



Post confirmation on the operational fesibility and impact on other parameters the optimal parameters were tried for a duaration of one week with a comprehensive control plan.

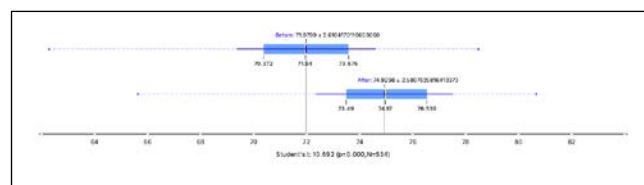
Figure 11



VII. CONCLUSION

After considering couple of combinations, the final one was recommended. Also, the model is being used to optimize the process during day-to-day operations.

The results before and after a given in Figure 12



Annexure 1 : List of Parameters Considered

Parameter	Type	Role
Date	Date	Skipped
Month	Date	Skipped
Quarter	Date	Skipped
RM Source	Categorical	Independent
Spray-1	Numeric	Independent
Spray-2	Numeric	Independent
Spray-3	Numeric	Independent
Coarse	Numeric	Skipped
Fine	Numeric	Skipped
Coarse TPH	Numeric	Independent
Fine TPH	Numeric	Independent
Hold Time	Numeric	Independent
Fill Temp	Numeric	Independent
Hold Temp	Numeric	Independent
Dosing Rate	Categorical	Independent
A Feed	Categorical	Independent
Yield	Numeric	Independent
Yield	Categorical	Independent

Annexure 2 : Feature Statistics

Name	Distribution	Mean	Median	Dispersion	Min.	Max.	Missing
N Spray-2		8.95537	8	0.401351	2.33333	24.6667	0 (0%)
N Spray-1		14.0521	13	0.361773	5	34.6667	0 (0%)
N Spray-3		2.31077	2.275	0.343506	0.213333	5.52	0 (0%)
N Fine TPH		94.3154	94.1875	0.207957	16.4583	199.923	0 (0%)
N Coarse TPH		257.021	259.375	0.159128	67.9167	384.667	0 (0%)
N Hold Time		42.6091	42.16	0.102861	29.79	55.5	0 (0%)
N Yield (t)		72.5931	72.475	0.039483	62.272	80.66	0 (0%)
N Hold Temp		56.2517	55.7	0.0365892	46.8	61.84	0 (0%)
N Fill Temp		78.7775	76.75	0.0188704	74.9	84.233	0 (0%)
G RM Source			RM2	1.1			0 (0%)
A Feed							0 (0%)

