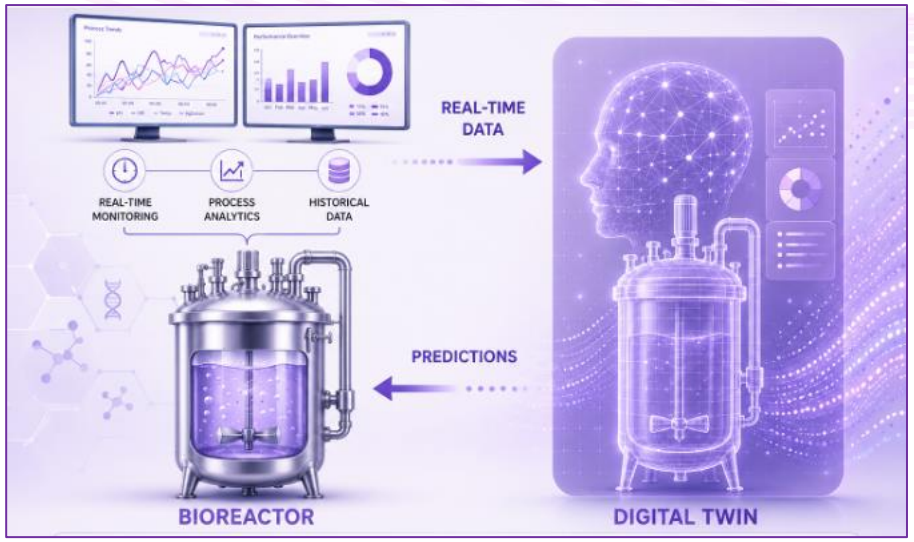
2026 DISCOVERY SUMMIT

EXPLORING DATA
INSPIRING INNOVATION



Democratising AI for Pharmaceutical Process Optimisation using JMP

Anurag Seksaria

(2026-US-30MP-2984)

PROCESS OPTIMIZATION | IMPROVED PRODUCT QUALITY | REDUCED RISK | HIGHER EFFICIENCY

Objective

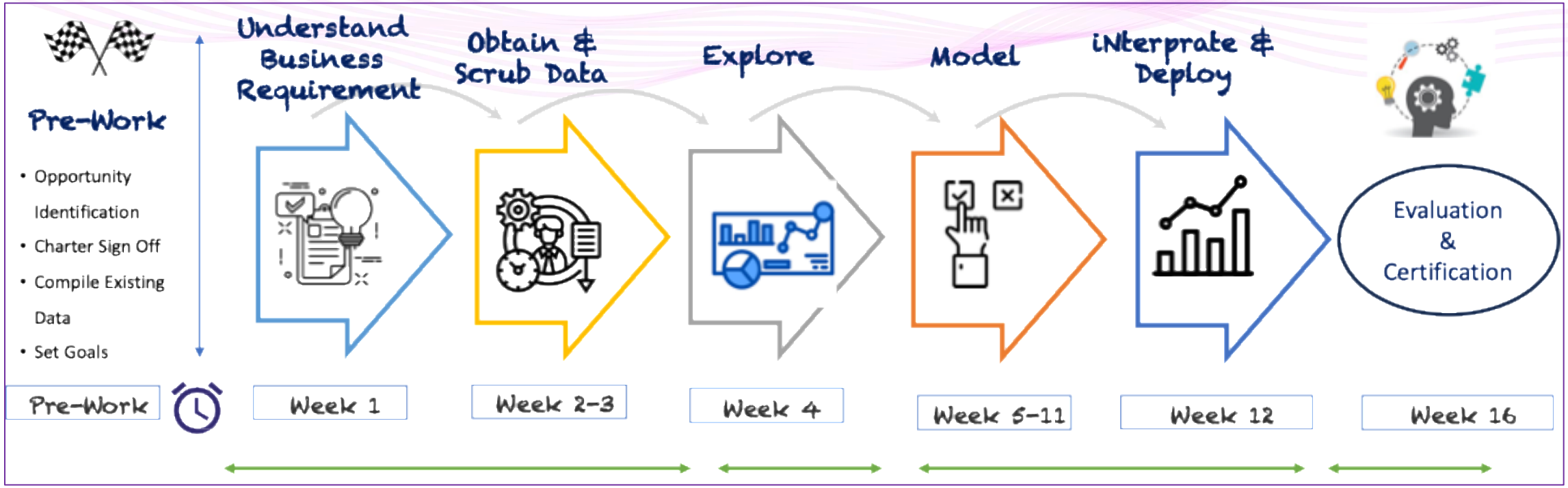
1. Optimisation of Yield for a Complex Pharma Process
2. Provide an approach which can be Democratised

Key Challenges

- Process parameters are manged ***"Real Time"*** while Yield is known only at the **"End of Fermentation"**
- High variability in Raw Material Quality difficult to assess

U-OSEMN Model

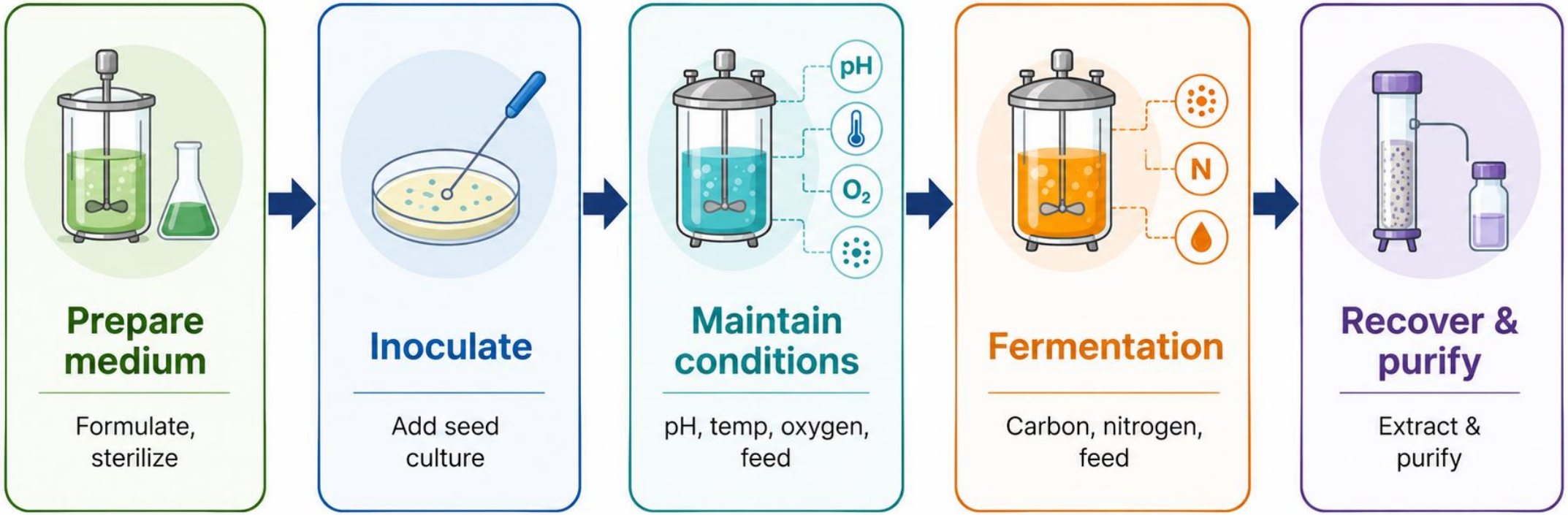
2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION



"A Taxonomy of Data Science" Hilary Mason and Chris Wiggins,

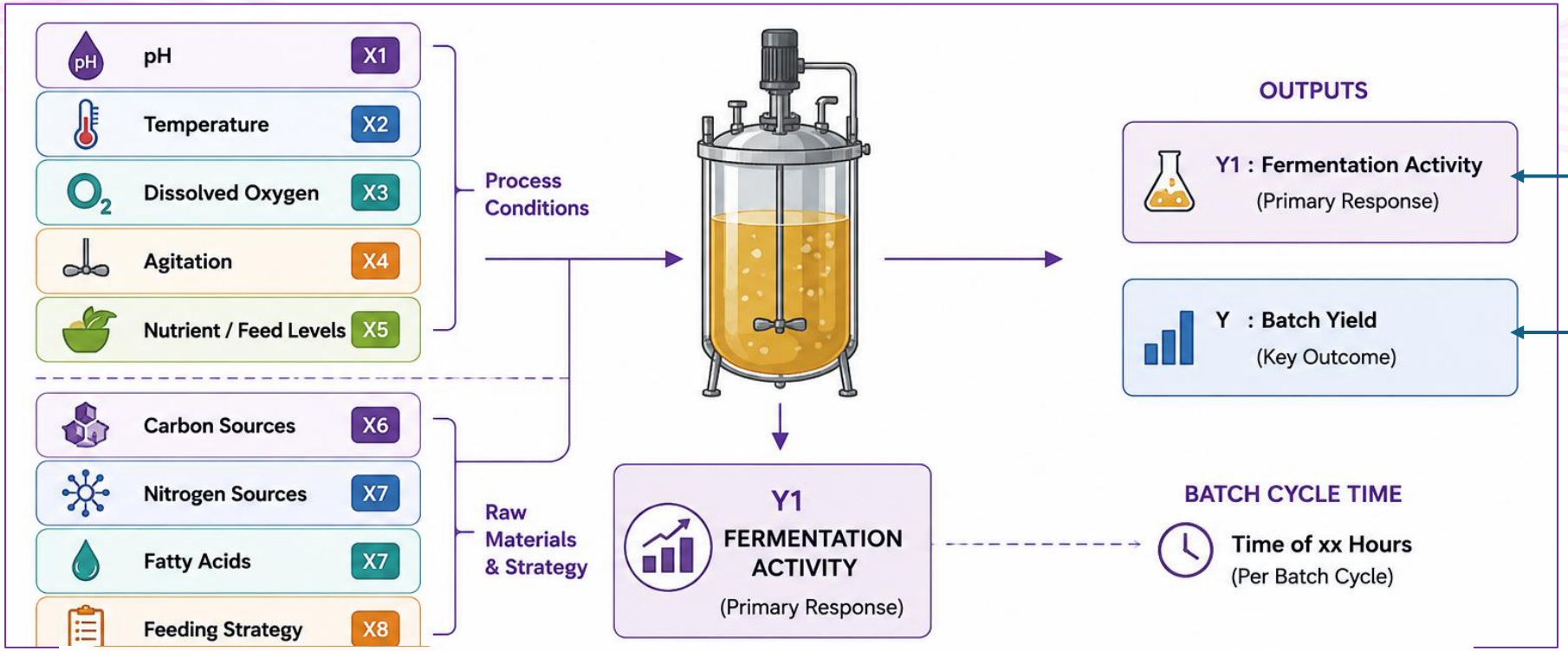
Controlled Microbial Fermentation Process

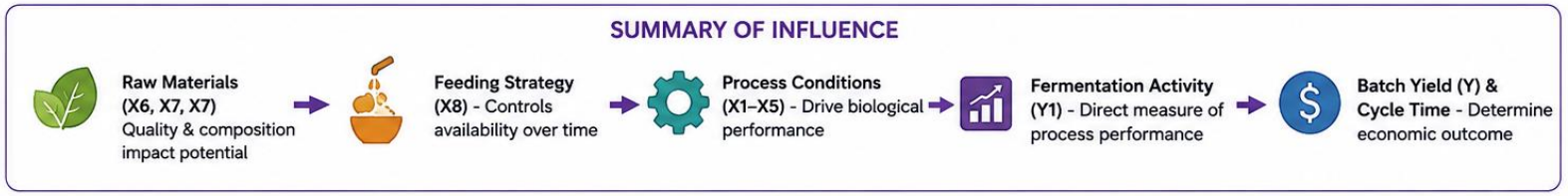
2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION



Influencing Factors

2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION

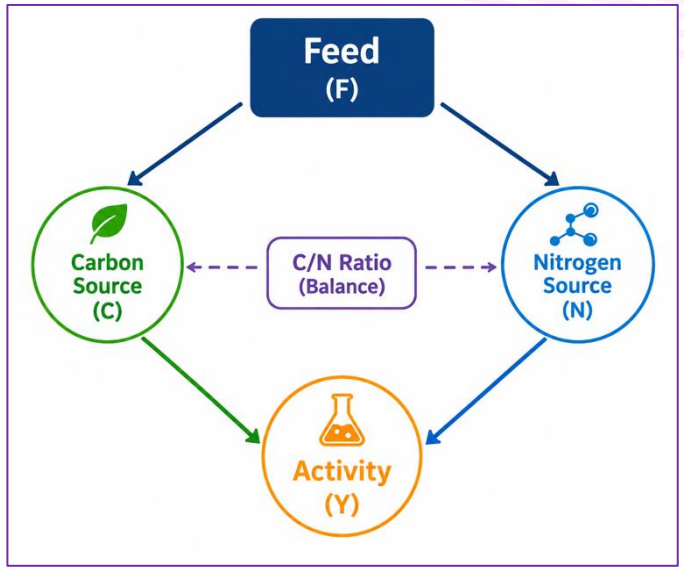




Interrelationship Between Factors

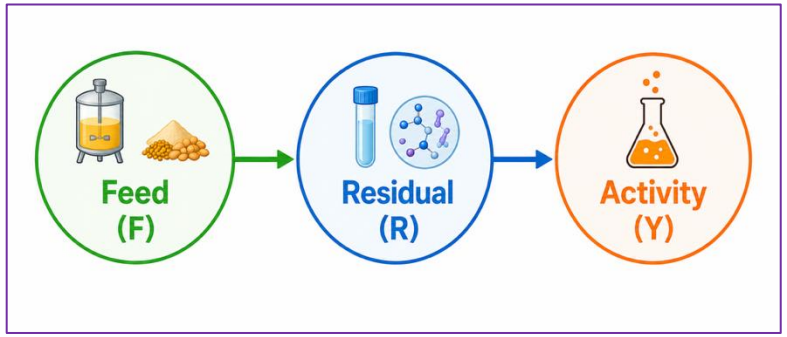
2026 DISCOVERY SUMMIT
EXPLORING DATA INSPIRING INNOVATION

Confounding



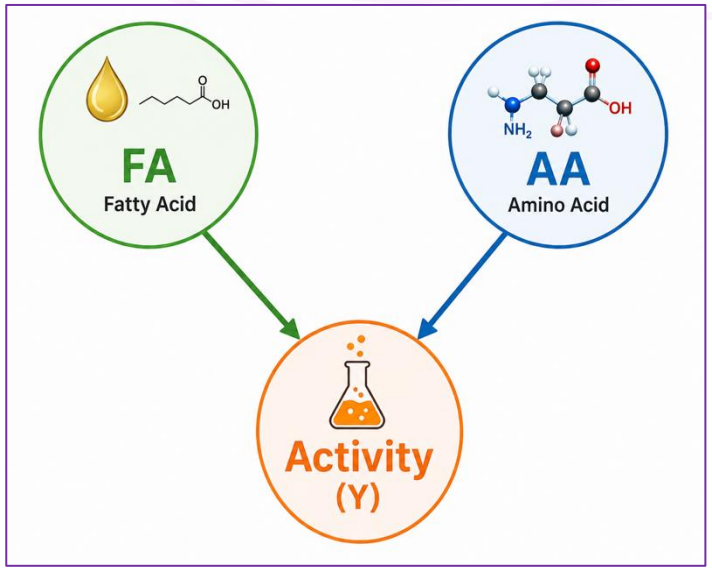
Feed -> Carbon & Nitrogen Carbon/ Nitrogen Ratio

Sequential



Feed v/s Residual

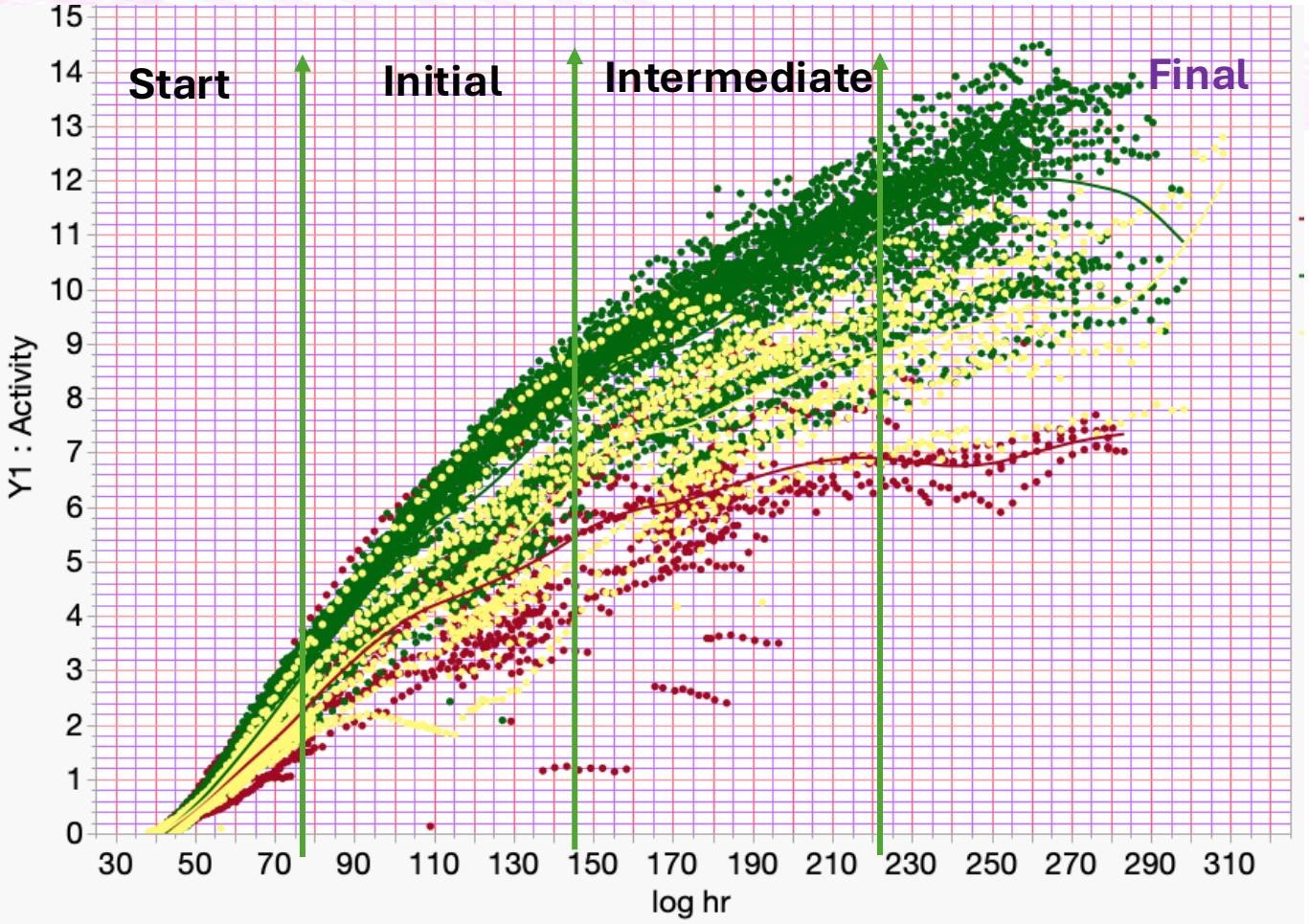
Converging



Fatty & Amino Acids independently impact Activity

Log Hours v/s Fermentation Activity (Y1)

2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION



Green : High Yield
Yello : Medium Yield
Red : Low Yield

For better analysis, the process was **SPLIT** into **4 STAGES** using "***Rate of Change"***

- Start
- Initial
- Intermediate
- Final

Bases for Creating SPLIT

2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION

Stage Start/High Yield

Summary of Fit

RSquare	0.834279
RSquare Adj	0.834141
Root Mean Square Error	1.064479
Mean of Response	3.897046
Observations (or Sum Wgts)	1206

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	-0.660512	0.066079	-10.00	<.0001*	-0.790155	-0.530868
log hr	0.0318449	0.000409	77.85	<.0001*	0.0310424	0.0326474

Stage Final/High Yield

Summary of Fit

RSquare	0.947313
RSquare Adj	0.947279
Root Mean Square Error	0.812768
Mean of Response	5.359206
Observations (or Sum Wgts)	1548

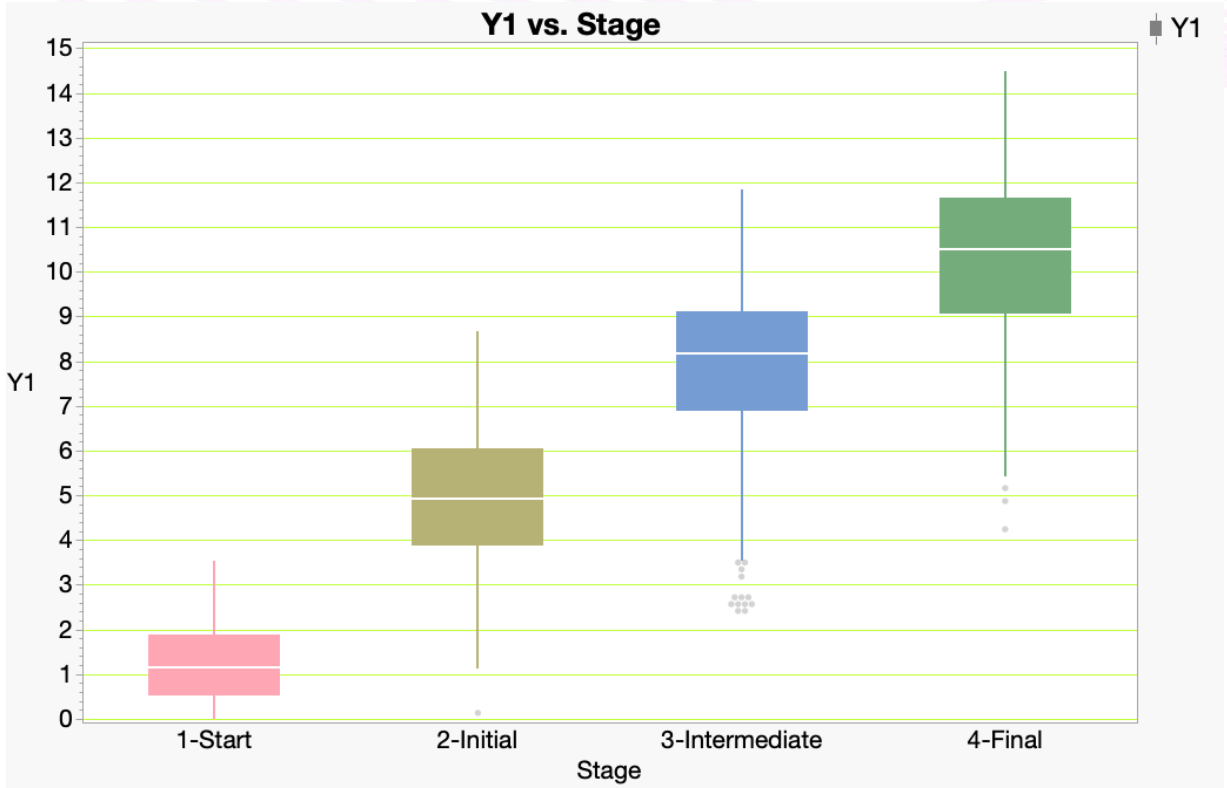
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	-1.170447	0.044278	-26.43	<.0001*	-1.257299	-1.083594
log hr	0.0450844	0.00027	166.72	<.0001*	0.044554	0.0456148

Regression Analysis by Low & High Yield Batches

Stage wise Activity

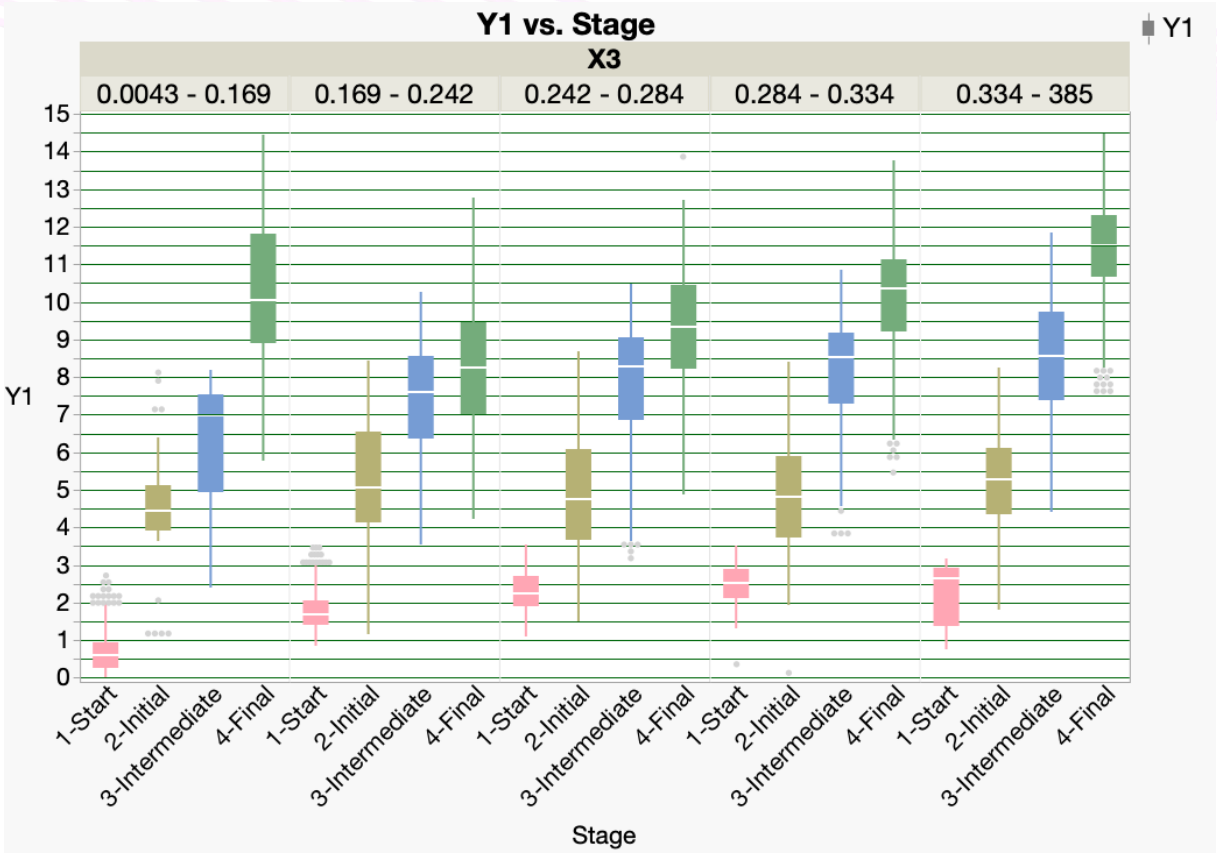
2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION



Significant difference between stages for both ***Averages*** and ***Variation*** for Activity

Exploration - Each Variable Stage Wise

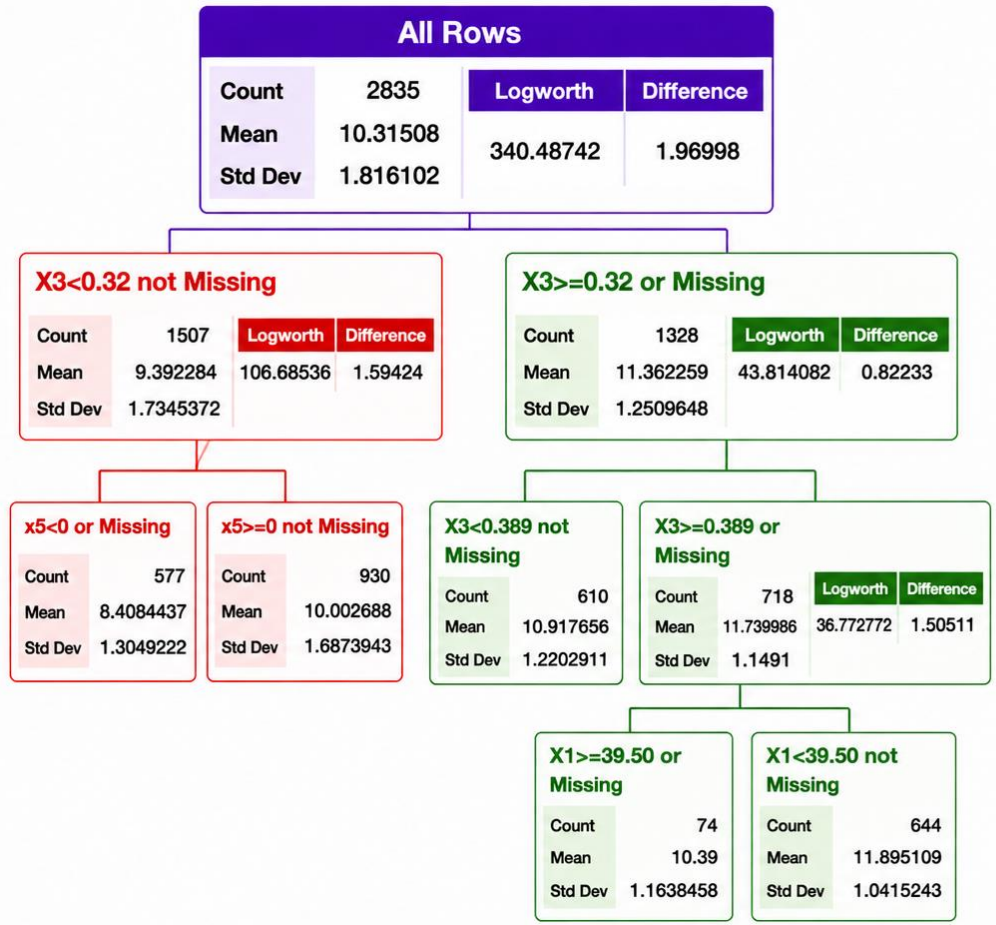
2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION

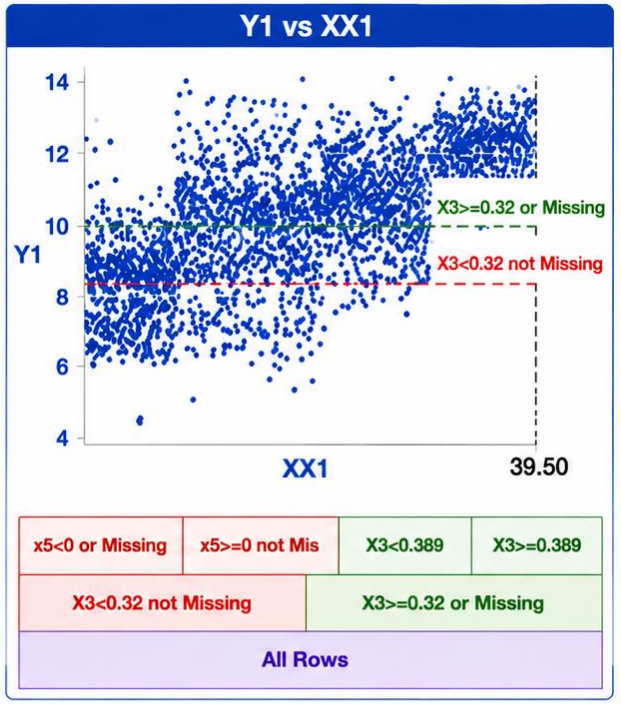


Variable X3 – The optimal levels vary with stages of fermentation

Tree – Stage Wise Optimization

2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION





RSquare	RASE	N	Number of Splits	AICc
0.430	1.3710569	2835	4	9846.76

11

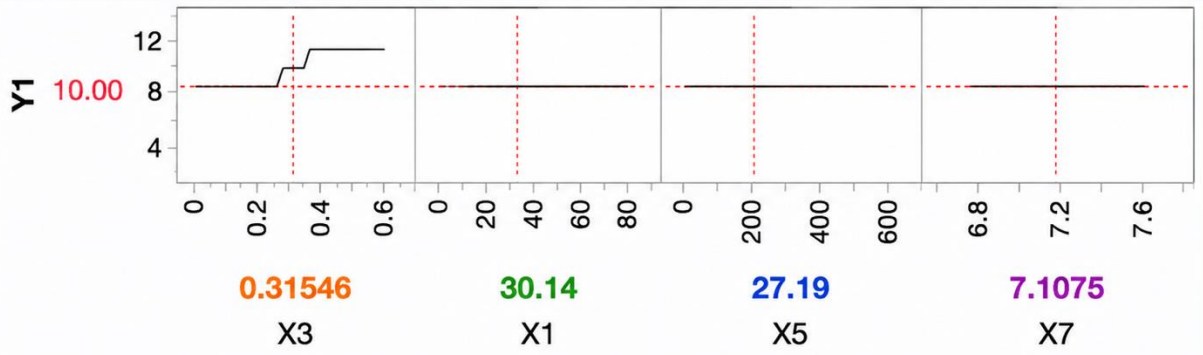
Fine Tunning Optimal with Profiler

2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION

Column Contributions

Term	Number of Splits	SS	Portion
X3	2	3066286.12	0.6214
X1	6	1089294.67	0.2208
X5	3	431444.028	0.0874
X7	3	347145.47	0.0704

Prediction Profiler



Fine Tuning between recommendations through TREE Analysis & Profiler

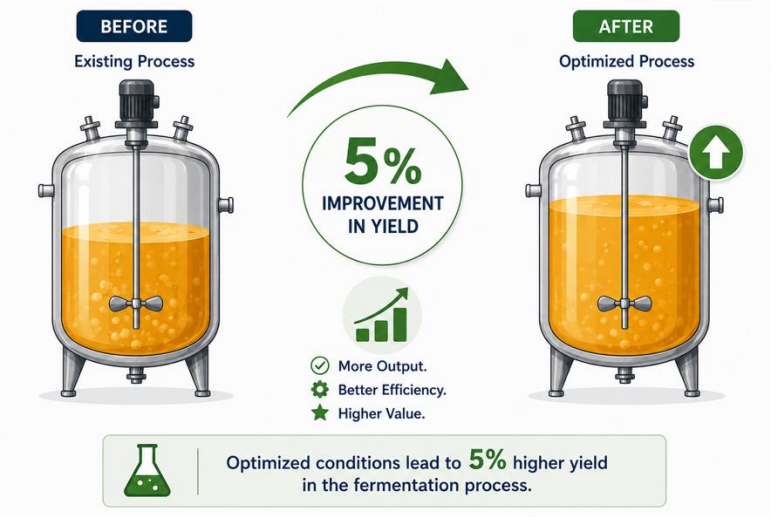
Bayesian Approach

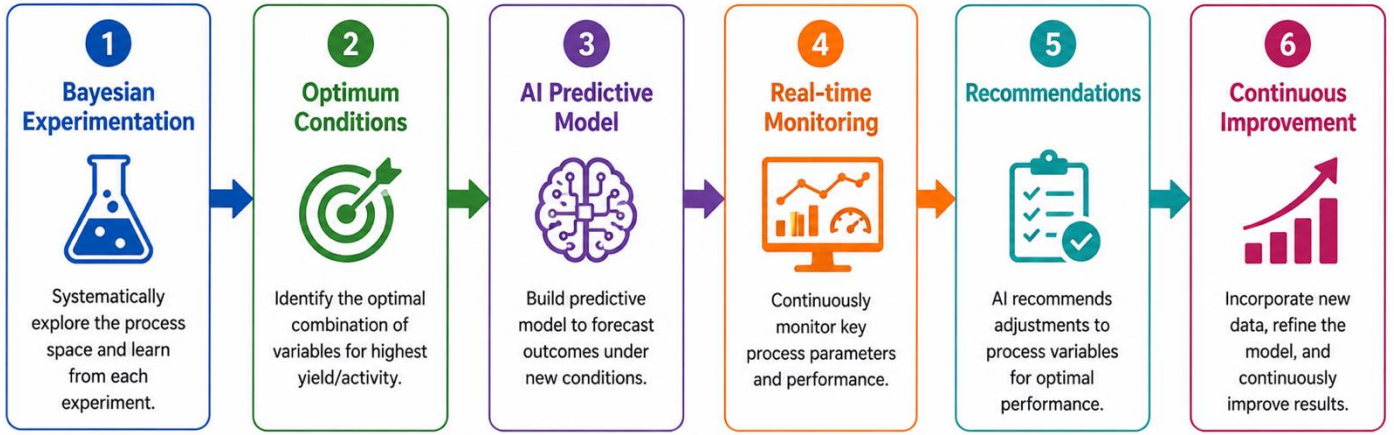
2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION

SN	ANALYSIS	RELATIONSHIP / OUTPUT	PURPOSE
1	Feed Duration Effect	Response = f(Feed Hours)	Determine optimum feed duration for maximum response & efficiency.
2	Input A–Feed Relationship	Response (Y) = f{Input A (y_1), Feed Rate (x_1)}	Understand combined effect of input A & feed rate on response.
3	Input B–Feed Relationship	Response (Y) = f{Input B (y_2), Feed Rate (x_1)}	Understand combined effect of input B & feed rate on response.
4	Input A–Additive 1 Relationship	Response (Y) = f{Input A (y_1), Additive 1 (x_2)}	Assess the effect of additive 1 addition with input A on response.
5	Input Relationship (A vs B with Additive 1)	Input A (y_1) = f{Input B (y_2), Additive 1 (x_2)}	Understand interrelationship between inputs & additive 1.
6	Additive Usage Year-on-Year Comparison	Year 1 vs. Year 2 (w.r.t. Additive 1, Additive 2)	Compare performance & consumption patterns of additive 1 & additive 2.
7	Additive 1 Level Impact	Response (Y) = f(Additive 1 Level)	Identify optimum level/range of additive 1 for desired response.
8	Additive 2 Level Impact	Response (Y) = f(Additive 2 Level)	Identify optimum level/range of additive 2 for desired response.
9	Process Interaction Key Inputs & Additives)	Response (Y) = f{Input A, Input B, Additive 1,2..., Feed Rate}	Evaluate interaction effects among key inputs, additives & feed rate.
10	Process Profile Analysis Over Time	Parameter 1 & Parameter 2 vs. Log Hours	Analyse trajectories of key parameters & process stability over time.
11	Correlation & Association Analysis	Correlation among Key Inputs, Additives, Response	Identify strong associations influencing the response & detect multicollinearity.
12	EOF Analysis	EOF by Time (using calculation)	Determine End of Process (EOF) at the right time.
13	Optimization Analysis	Multi-response optimization using key inputs, additives & feed rate	Identify optimal combination of factors for maximizing response & minimizing cost.
14	Outlier & Anomaly Detection	Detection of outliers in Key Parameters & Response	Ensure data quality & reliability by identifying abnormal observations.
15	Sensitivity Analysis	Impact of ± change in inputs/additives on response	Identify critical factors & robustness of the process.
16	Contribution / Pareto Analysis	% Contribution of each factor to response variation	Focus improvement efforts on factors that contribute the most to performance.
17	Robustness & Stability Assessment	Process Performance across batches / days / conditions	Validate consistency & robustness of process under normal operating ranges.
18	Predictive Modeling (AI/ML)	Build predictive model for response using key factors	Predict future outcomes, support decision making & enable real-time process control.

Results & Deployment

2026 DISCOVERY SUMMIT
EXPLORING DATA
INSPIRING INNOVATION





Digital Twin Helps Operations Team to Run Process Optimally


2026
DISCOVERY
SUMMIT
EXPLORING DATA
INSPIRING INNOVATION