Machine learning platform to assess materials performances after Gamma and X-ray sterilization

SARTORIUS



With the support of:

Introduction

From Sterilization Challenges to AI Prediction



Is it possible to predict the performance of polymers using A.I ?

SARTORIUS

Global leader in the design and manufacture of single-use systems (SUS)

STERILIZATION

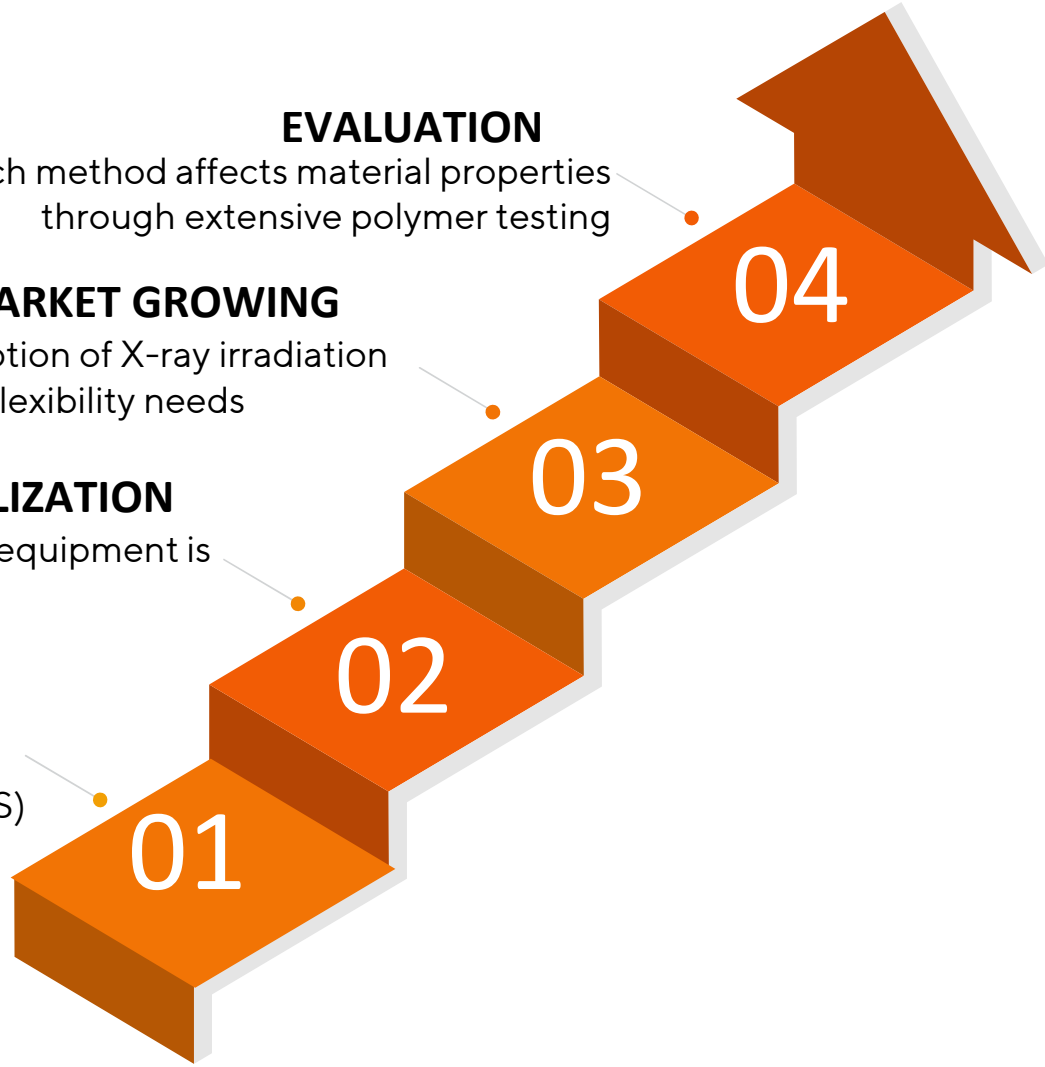
Using polymer materials in Sartorius equipment is a critical step in ensuring safety

MARKET GROWING

Leads to broader adoption of X-ray irradiation to meet capacity and flexibility needs

EVALUATION

How each method affects material properties through extensive polymer testing



Why Sterilization Equivalency Is Difficult to Demonstrate ?

Supplementing Gamma with X-ray

Gamma implementation x-ray require a science-based rationale to accept X-ray as an equivalent technologies.

Market requires proving that X-ray does not negatively impact polymer performance compared to Gamma.



Physical testing is a bottleneck

Standard validation relies on ageing studies (12 to 36 months).

Re-testing every material variation for X-ray creates multi-year delays and high experimental costs.



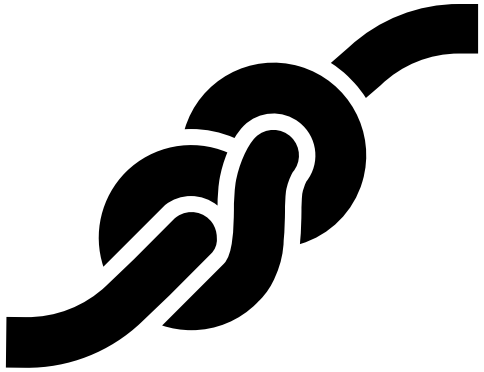
Data gap

Gamma: Deep historical data available (30+ years of characterization).

X-ray: Data is currently sparse and expensive to generate.



Towards Predictive Polymer Engineering



01 Prediction of polymer performance is achieved through machine learning

02 Model generation and validation performed in collaboration with **Quaisr**

Phase 1 Quaísr

03 Can we define the minimum testing required to run the ML models?

Phase 2 Quaísr

Phase 1: ongoing | Phase 2: Target end 2026

Machine Learning For Polymer Property Prediction

What is Machine Leaning(ML) ?

Machine Learning

What is ML ? 1

Machine Learning is the study of methods for programming computers to learn[1].



Type of ML ? 2

- Unsupervised learning
- Semi-supervised
- Supervised learning

Advantages ? 3

- Solving problems via statistical inference
- Reducing the constraints over time

How ? 4

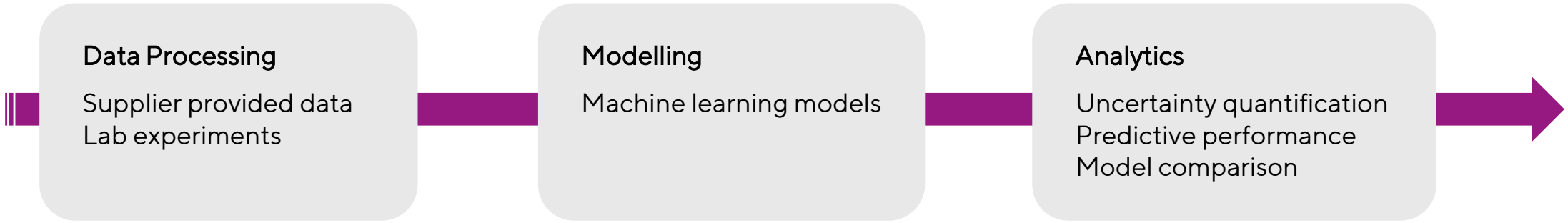
Communicate with machine : python, pandas, numpy, pytorch, sklearn...etc

NumPy PyTorch scikit learn Machine Learning Using Python TensorFlow

[1] : "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." — Tom M. Mitchell (1997)

SARTORIUS

Advanced machine learning enables polymer property prediction



Predict Polymer Properties

Data Processing

Supplier provided data Lab experiments

...with targeted data

Process new experimental data on demand

Improve modelling

Leverage cutting edge data-driven-modelling

Thrive on probabilistic data

Supplier provided data Lab experiments

Supervised machine learning methods learn from a data of input/output pairs

Bayesian regression

Uncertainty quantification

Assumes linear/polynomial relation between inputs and targets

Decision trees

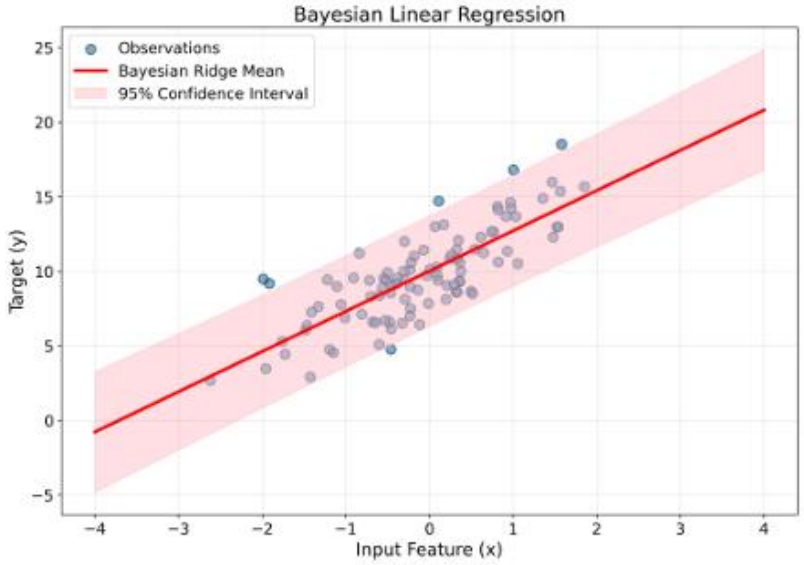
No uncertainty quantification

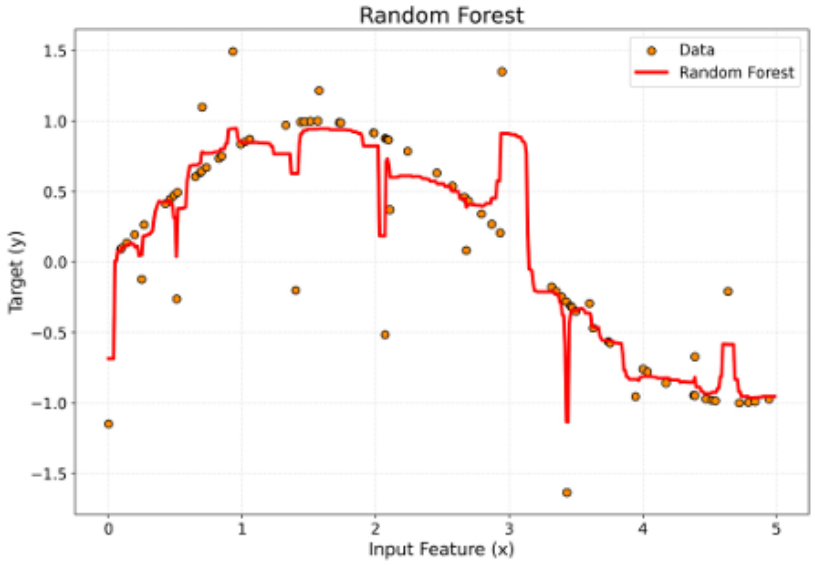
Fits general input and output relations

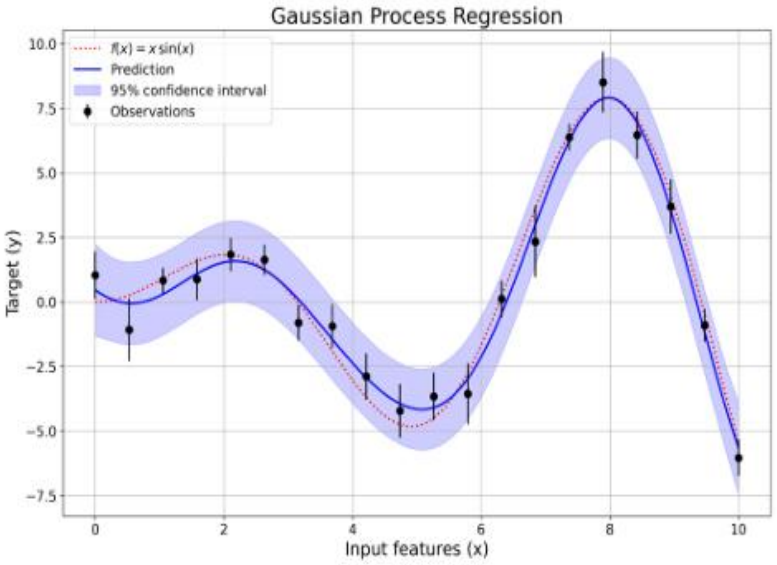
Gaussian processes

Uncertainty quantification

Fits general input and output relations

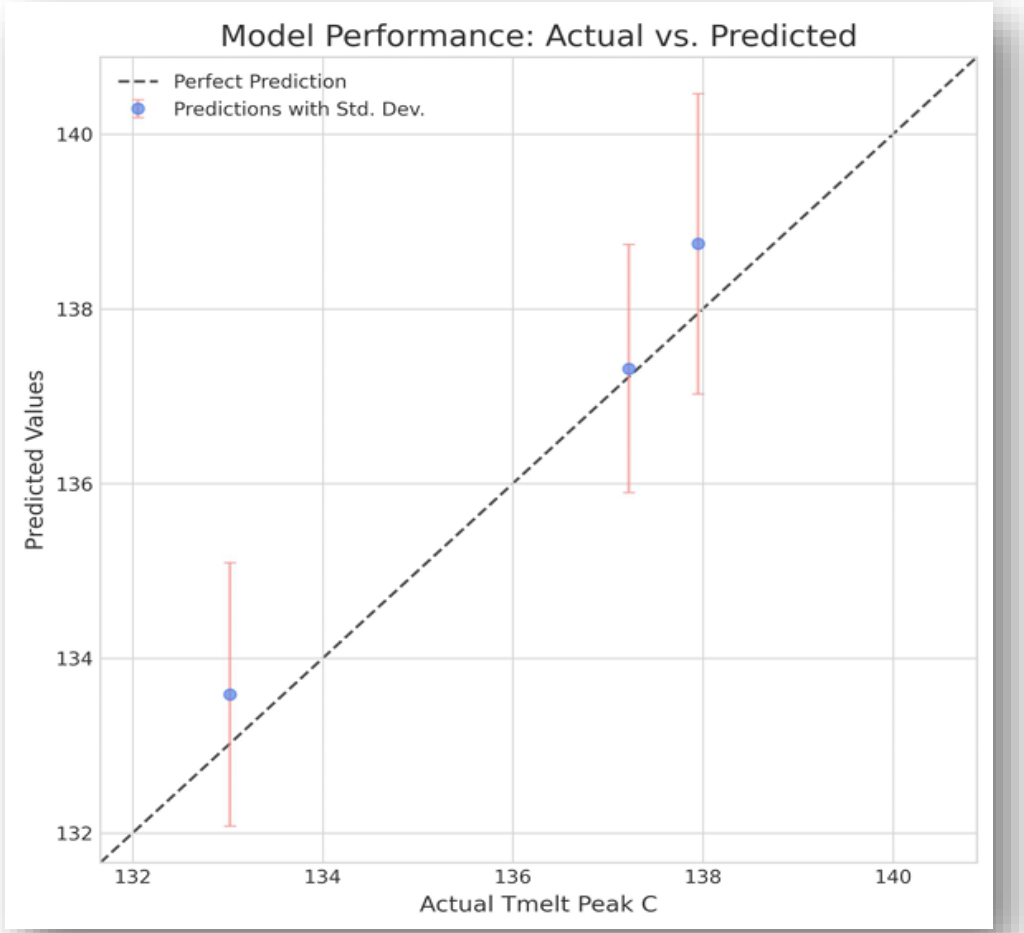






Measures Of Performance : Parity Plot & Key Performance Indicators

- Model validation and performance are evaluated by keeping a subset of data aside from model training. That data is used for computation of performance metrics.



Parity plot

Measured values plotted against model predictions. Model predictive performance is good when the error bars of the points go through the bisector line

R^2 Score

Measure of how well observed data are replicated by the model

Root mean squared error

Average distance of the data from the model

Gaussian Process Approach: Principles and Strengths

- Offer a robust Bayesian method for regression analysis, classification and uncertainty quantification

Let the data speak for itself

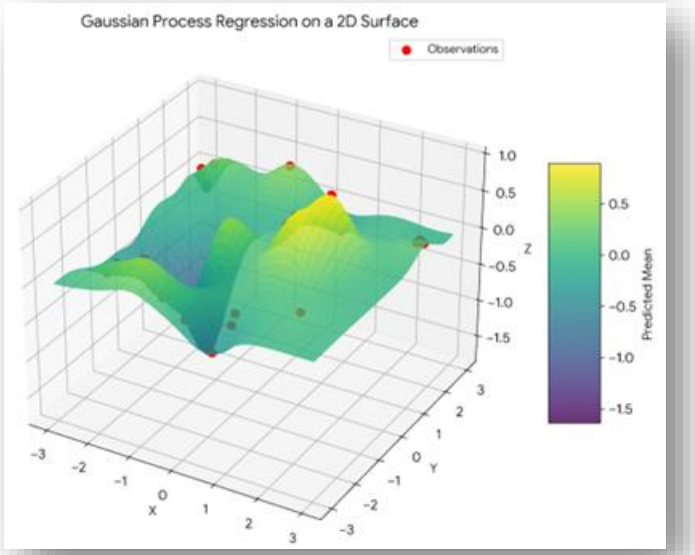
Make little to no assumptions about the underlying shape of parameter dependencies.

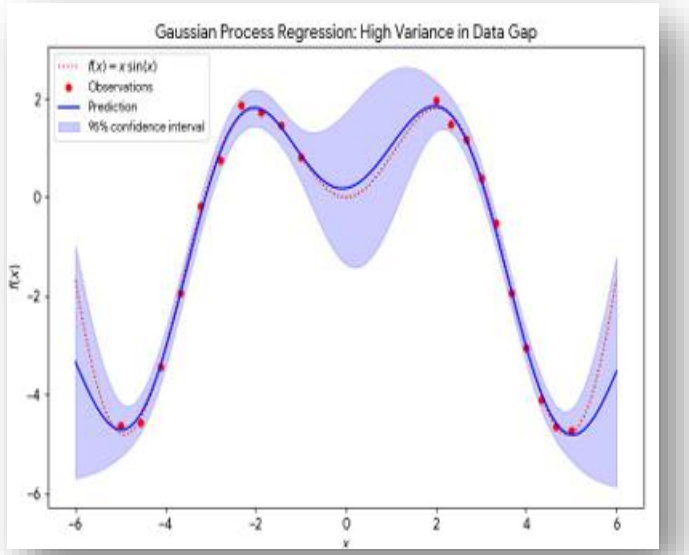
Quantify uncertainty

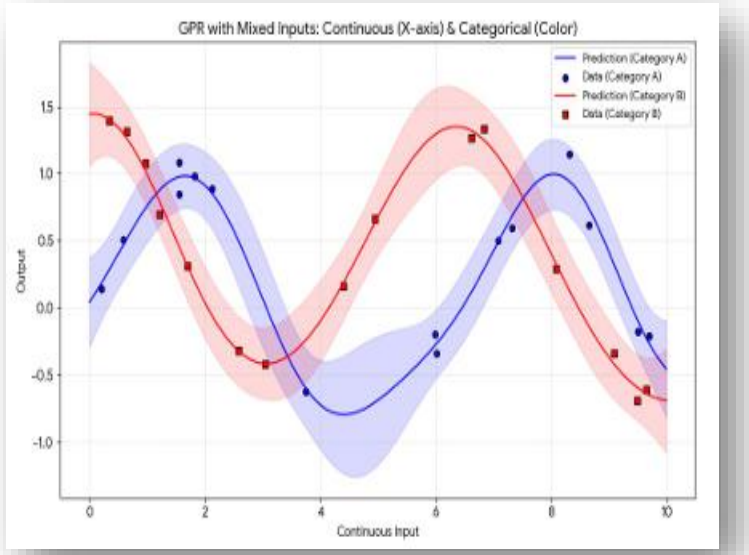
Provide uncertainty estimates in areas where we have limited or no data.

Mixed data

Mix categorical and continuous parameters.







Method works well for small datasets, hybrid categorical and numerical data, large number of input features.

SARTORIUS

Data Framework for Predictive Modeling

SARTORIUS

Holistic three-level testing approach[1,2]

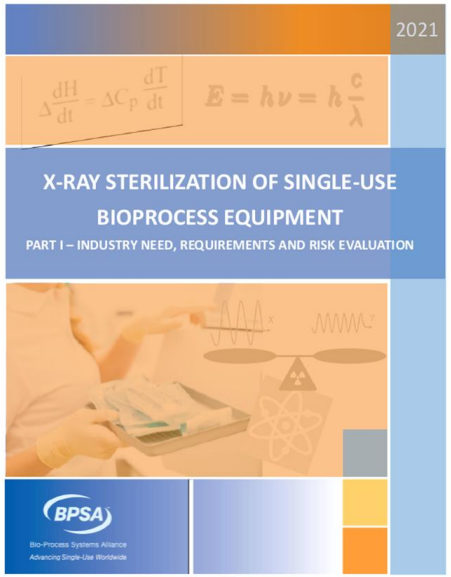


Quality Test Matrices

- ✓ Materials Tests*
- ✓ Physical Tests
- ✓ Functional Tests

Connectors & Valves | Containers & Film | Sensors | Tubing | Filters | ✓ Assemblies

Materials Assessment
Color, DSC*, TGA*, FTIR*



Polymers	DSC	FTIR	DMA	Tensile test	Radicals	Colorimetry
PP	✓	✓	✓	✓	✓	✓
HDPE	✓	✓	✓	✓	✓	✓
LDPE	✓	✓	✓	✓	✓	✓

- Data generated by single use system (SUS) after gamma and X-ray irradiation

- 1N.Dupuy, S.R.A. Marque, L.S. Fifield, M.Pharr, D. Staack, S. D. Pillai, L.Nichols, M.K. Murphy, and S.Dorey, Supplementing Gamma Sterilization with X-Ray and E-Beam Technologies, An International Industry and Academia Collaboration, BioProcessInternational, 20(3), 2022, 24-28
- 2BPSA,X-Ray Sterilization of Single-Use BioProcessEquipment, Part 1: Industry Need, Requirements & Risk Evaluation, 2021

Resine suppliers data

- Data form resine suppliers : avaible in TDS before gamma and X-ray irradiation

	Polymer type	Density(g/cm3)	Test method	Tensile modulus(Mpa)	Test Method	tensile stress(Mpa)	Test method	Point Vicat(°C)	Test Method	Melting temperature(°C)	Test method
HDPE 1	Butene Copolymer	0,96	ISO 1872-1	1500	ISO 527-2/1B	31	ISO 527-2/1B	/	/	250	n.a
HDPE 2	Ethene-Butene copolymer	0,963	ISO 1183	1350	ISO 527	30	ISO 527	71	ISO 306	230	n.a
HDPE 3	Ethene Butene copolymer	0,952	ISO 1183	1000	ISO 527-1, 2	24	ISO 527-1, 2	64	ISO 306	230	n.a
HDPE 4	Butene Copolymer	0,953	ISO 1872	1150	ISO 527 -1, 2	26	ISO 527 -1, 2	/	/	/	/
HDPE 5	Hexene copolymer	0,953	ASTM D1505	1270	ASTM D790	28	ASTM D638	125	ASTM D1525	/	/
HDPE 6	hexene copolymer	0,954	ISO 1183	1378	D 790	27,6	D 638	/	/	264	D 3417
HDPE 7	butene copolymer	0,962	D 792	1280	D 790	29	D 638	129	D 1525	/	/
HDPE 8	Hexene copolymer	0,947	ISO 1183	800	DIN EN ISO 527	21	DIN EN ISO 527	73,1	ISO 306	/	/

Data Pretraetement

Gaussian process implementation enables training with missing data

Conventionally the materials without the Point Vicat value would need to be removed when fitting a model.

	Polymer type	Density(g/cm3)	Test method	Tensile modulus(Mpa)	Test Method	tensile stress(Mpa)	Test method	Point Vicat(°C)	Test Method	Melting temperature(°C)	Test method
HDPE 1	Butene Copolymer	0,96	ISO 1872-1	1500	ISO 527-2/1B	31	ISO 527-2/1B	/	/	250	n.a
HDPE 2	Ethene-Butene copolymer	0,963	ISO 1183	1350	ISO 527	30	ISO 527	71	ISO 306	230	n.a
HDPE 3	Ethene Butene copolymer	0,952	ISO 1183	1000	ISO 527-1, 2	24	ISO 527-1, 2	64	ISO 306	230	n.a
HDPE 4	Butene Copolymer	0,953	ISO 1872	1150	ISO 527 -1, 2	26	ISO 527 -1, 2	/	/	/	/
HDPE 5	Hexene copolymer	0,953	ASTM D1505	1270	ASTM D790	28	ASTM D638	125	ASTM D1525	/	/
HDPE 6	hexene copolymer	0,954	ISO 1183	1378	D 790	27,6	D 638	/	/	264	D 3417
HDPE 7	butene copolymer	0,962	D 792	1280	D 790	29	D 638	129	D 1525	/	/
HDPE 8	Hexene copolymer	0,947	ISO 1183	800	DIN EN ISO 527	21	DIN EN ISO 527	73,1	ISO 306	/	/

Suppliers missing data

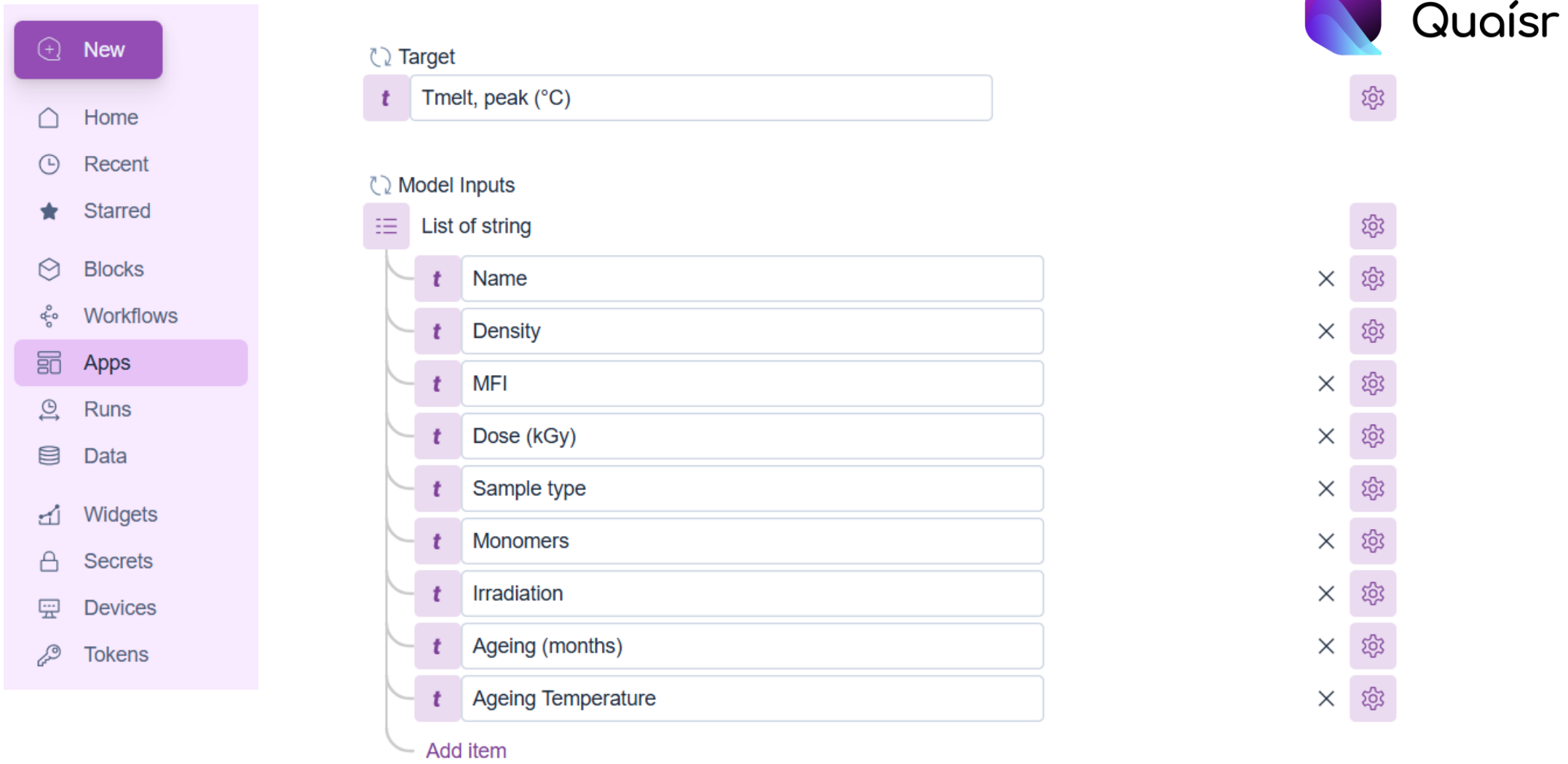
Implementation

- Gaussian process parameters fitted via stochastic gradient descent
- The missing Point Vicat value is sampled from a prior distribution at each Gaussian process fitting iteration.
- The Point Vicat prior distribution set based on a range of plausible Point Vicat values
  - Normal distribution with mean and standard deviation based on the values for other polymers.

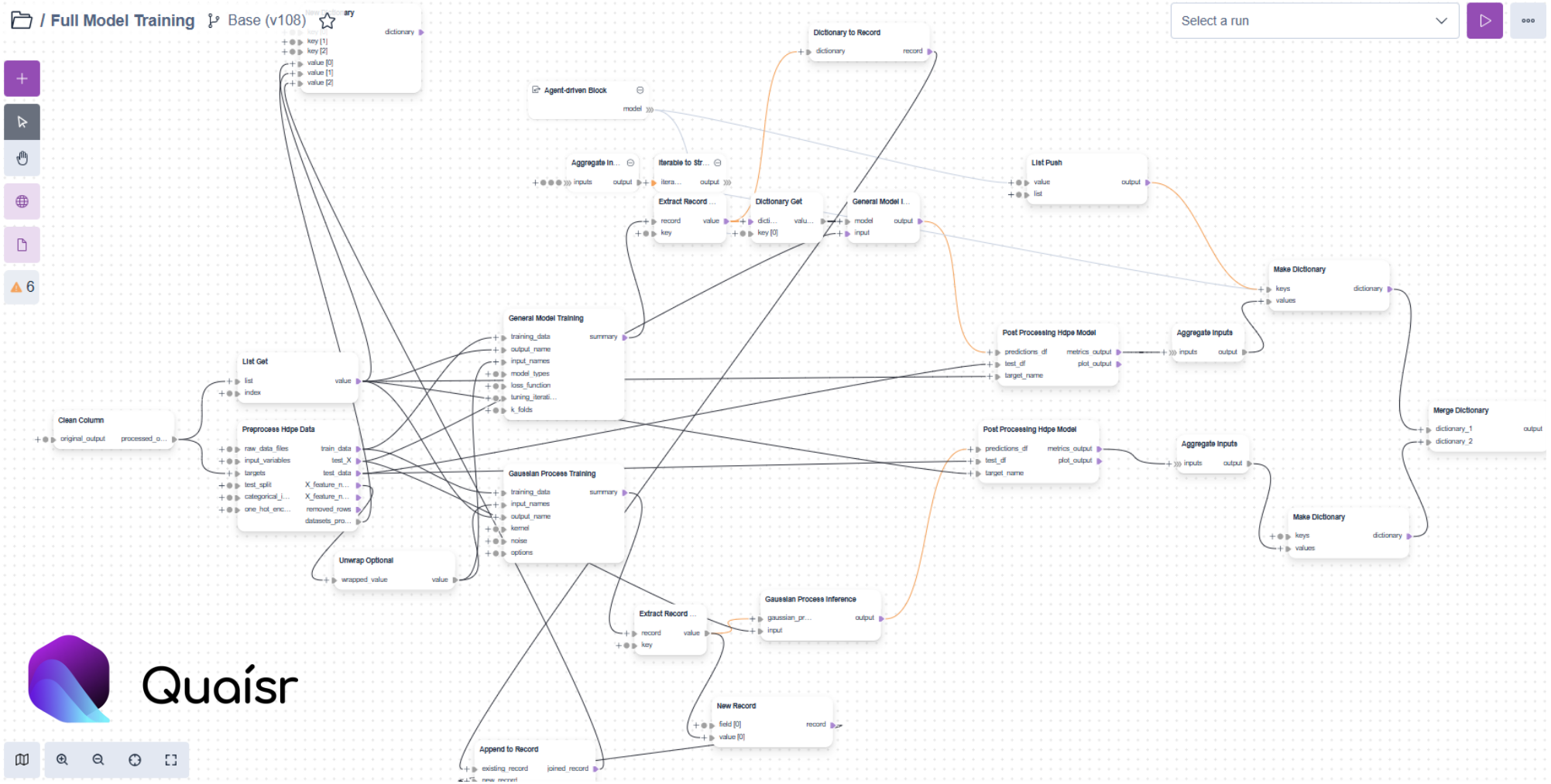
Machine Learning Models

SARTORIUS

Quaisr Platform Interface



Machine Learning workflow network

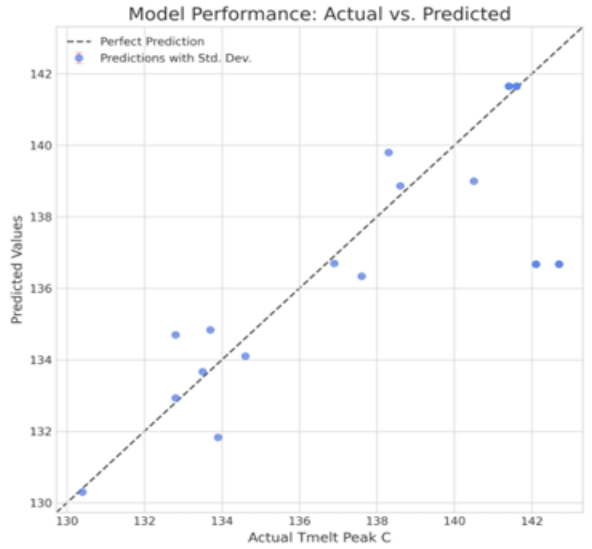


Restricted

Model Prediction Performance Based on Gaussian Process Regression

- predictive performance for Tmelting temperature improve with more data

HDPE 1148 Dataset



Outputs

metrics_output

Log output

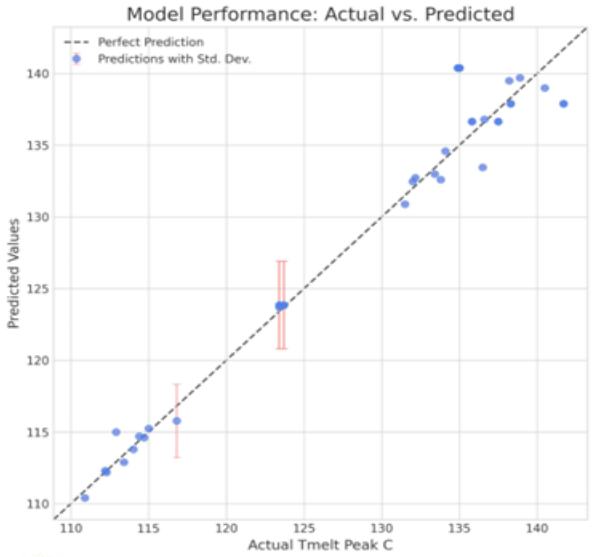
dictionary (3 items)

r2 0.531

mae 1.712

rmse 2.714

HDPE + LDPE 2794 Dataset



Outputs

metrics_output

Log output

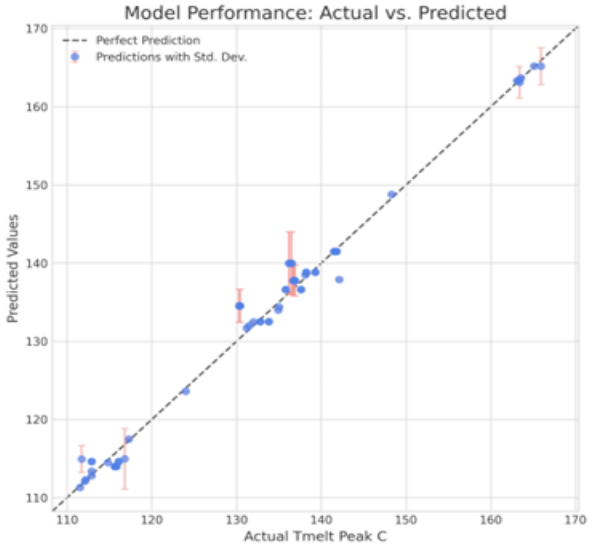
dictionary (3 items)

r2 0.956

mae 1.339

rmse 2.141

LDPE + HDPE + PP 3642 Dataset



Outputs

metrics_output

Log output

dictionary (3 items)

r2 0.985

mae 1.241

rmse 1.788

Conclusion

Conclusion

- The modeling requires a significant amount of data generated internally as well as from suppliers (before and after irradiation).
- The Gaussian Process approach appears to be the most appropriate considering our current dataset.
- Machine learning models have been developed to predict certain properties after irradiation and ageing.
- Predictions have been generated, and we are currently validating the models by testing them on new polymer grades.
- Validation with previously untested polymers should be carried out in Phase 2.

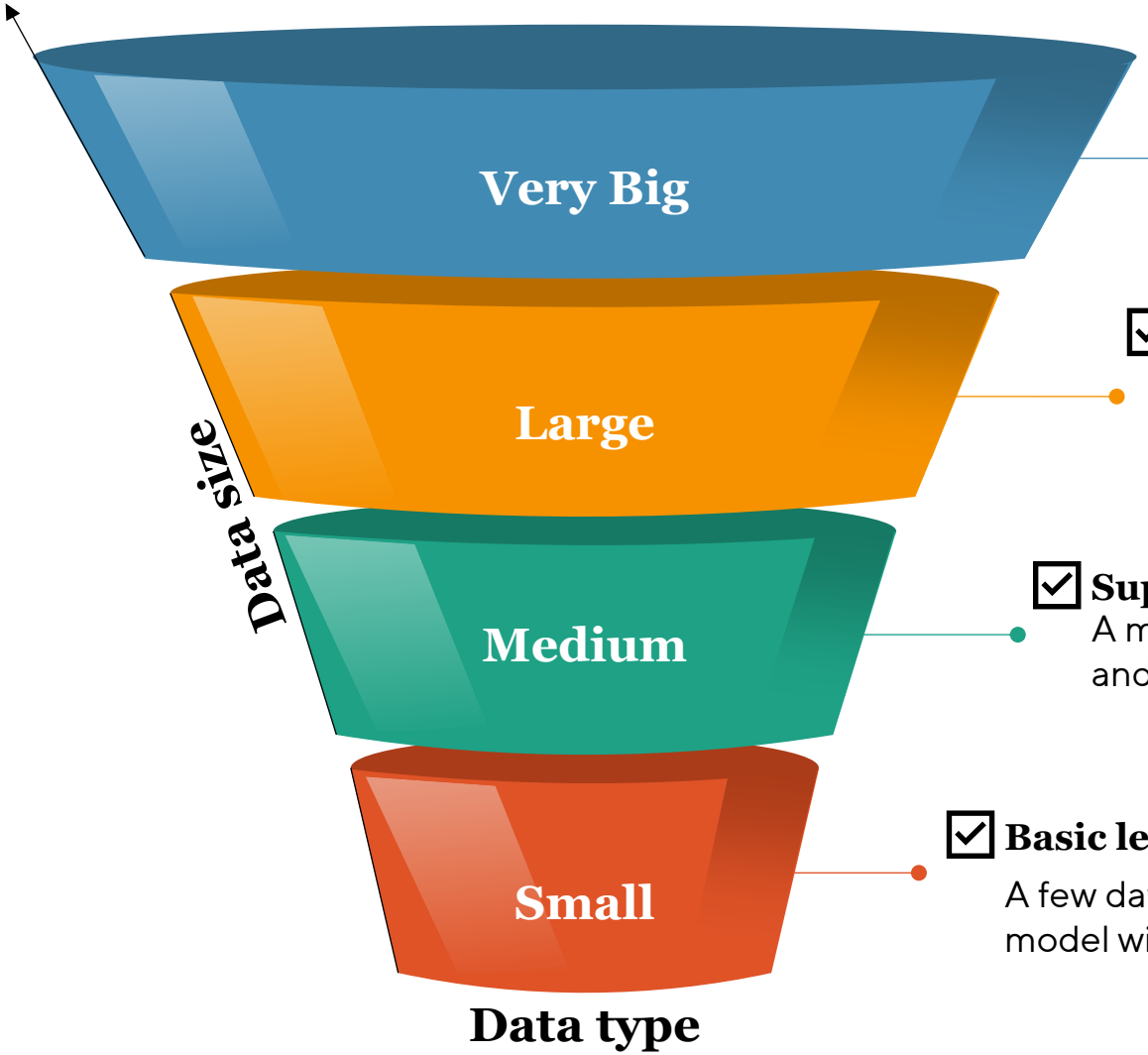
Thank you.

Dr. Samuel Dorey
Principal Scientist Materials & Irradiations
Mobile +33 6 37 18 11 43
Samuel.dorey@sartorius.com

BOUT Hakim
Materials Scientist
PD FMT – S & T
Hakim.bout@sartorius.com

SARTORIUS

Why not neural networks ? (appendix 1)



Neural learning
A model with a very large database, potentially thousands or millions of data points, to effectively train the model.

Unsupervised learning
A model with a large number of data points and multiple variables to build a robust model.

Supervised learning
A model with a moderate number of data points and more variables to start building the model.

Basic learning
A few data points are sufficient to build a simple model with a small number of variables.