Applications and Challenges of Machine Learning Applied to the Chemical Recovery Cycle

Dr. Andrew K. Jones

SUOMEN SOODAKATTILAYHDISTYS
FINNISH RECOVERY BOILER COMMITTEE

Outline

- What types of Real-time predictive models do we need?
- Specific examples of modeling applied to the Kraft recovery cycle
- Issues with data that must be considered
- Feature identification methods
- A Brief History

What Types of Real-Time Models Do We Need?

- Type 1: Forecasting Models predict future target values using past data and current conditions, aiding early detection of target deviations.
- Type 2: Prescriptive Models use only causal features as variables, assisting process adjustments or closed-loop control. They're less accurate than forecasting models but improve with better instrumentation. They're also useful for offline root cause analysis.

A List of Modeling Applications in the Kraft Recovery Cycle (all of these have been attempted)

Type 1 (Forecast)

- Fouling rate predictions – using SH Strain gages
- Reduction efficiency interpolation – between intermittent tests
- Reduction efficiency forecast – future values
- Residual carbonate interpolation in lime – in between intermittent tests
- Product solids from evaporators – in between intermittent tests
- Future values of boiler emissions.
- TTA and CE in white liquor – in between intermittent tests

Type 2 (Prescriptive)

- Reduction efficiency (with in-line measurements required)
- Superheater fouling root causes
- NOx reduction on a recovery boiler

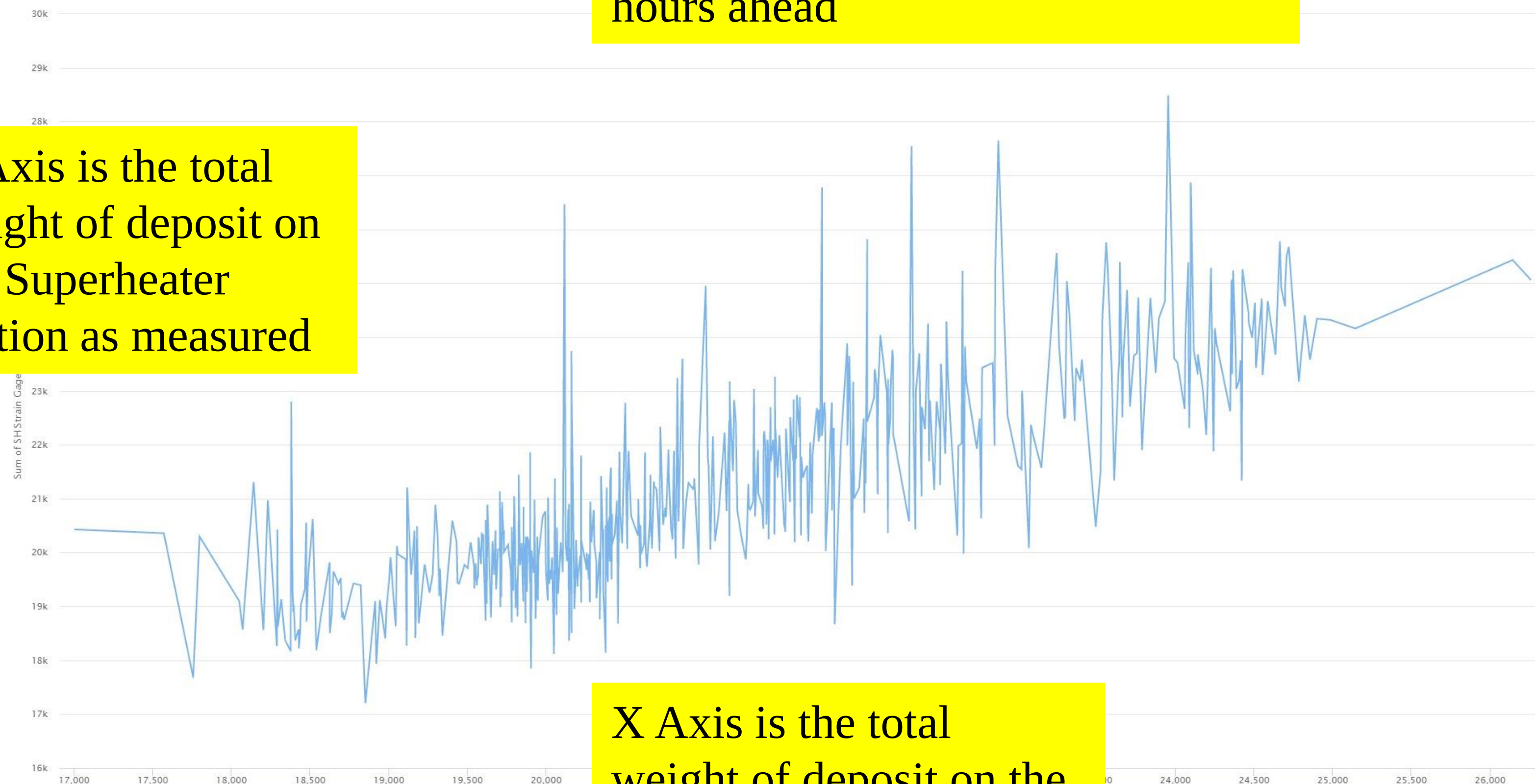
EXAMPLE 1 – FORECAST TYPE MODEL

In this case ONLY previous values of the target variable are used

Predicting Future Fouling

This ML technique uses Sliding Range validation instead of data partitioning. It predicts future target values using past data. It learns the impact of sootblowers on deposit weight based on frequency and location, aiding selection of optimal sootblowers.

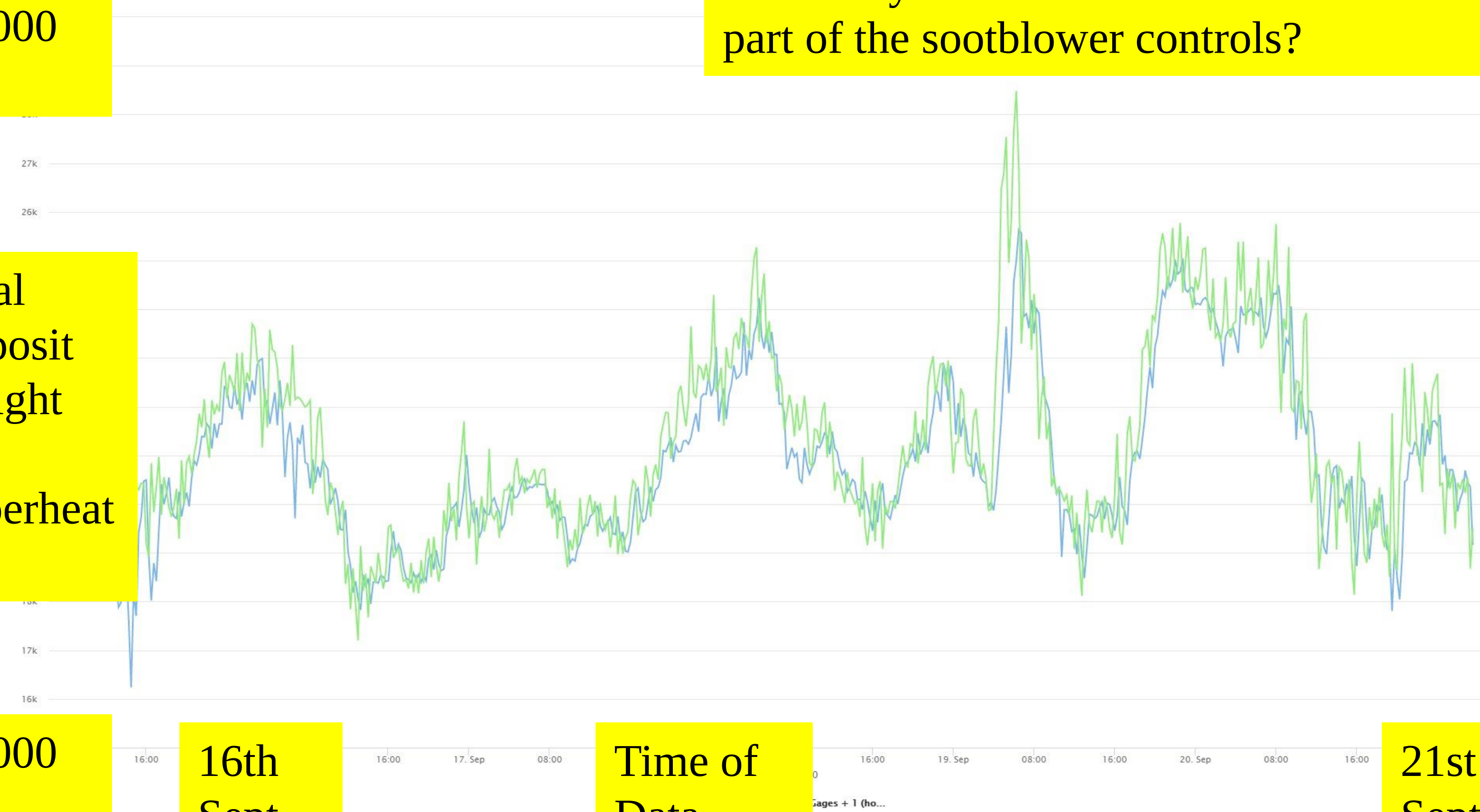
Actual versus Prediction Fouling 4 hours ahead



Y Axis is the total weight of deposit on the Superheater section as measured

X Axis is the total weight of deposit on the Superheater section as predicted by the 4 hour look-ahead model

Predicted one hour ahead (green) versus actual (blue) – excellent ability to forecast future fouling – I strongly suspect this is because the MI tool "learns" the pattern created by sootblowers – Can we use this as part of the sootblower controls?



EXAMPLE 2 – PRESCRIPTIVE TYPE MODEL

Method Used

The dataset includes process variables and strain gage weight changes. Process variables, manipulated hourly by operators, are analyzed along with weight deltas, typically focusing on hourly changes in superheater weight. Gradient-Boosted Trees are the most successful machine learning method used. Relationships are visualized using ALE (Accumulated Local Effects) plots, derived from solved Gradient Boosted Trees.

Model Performance and VIPs

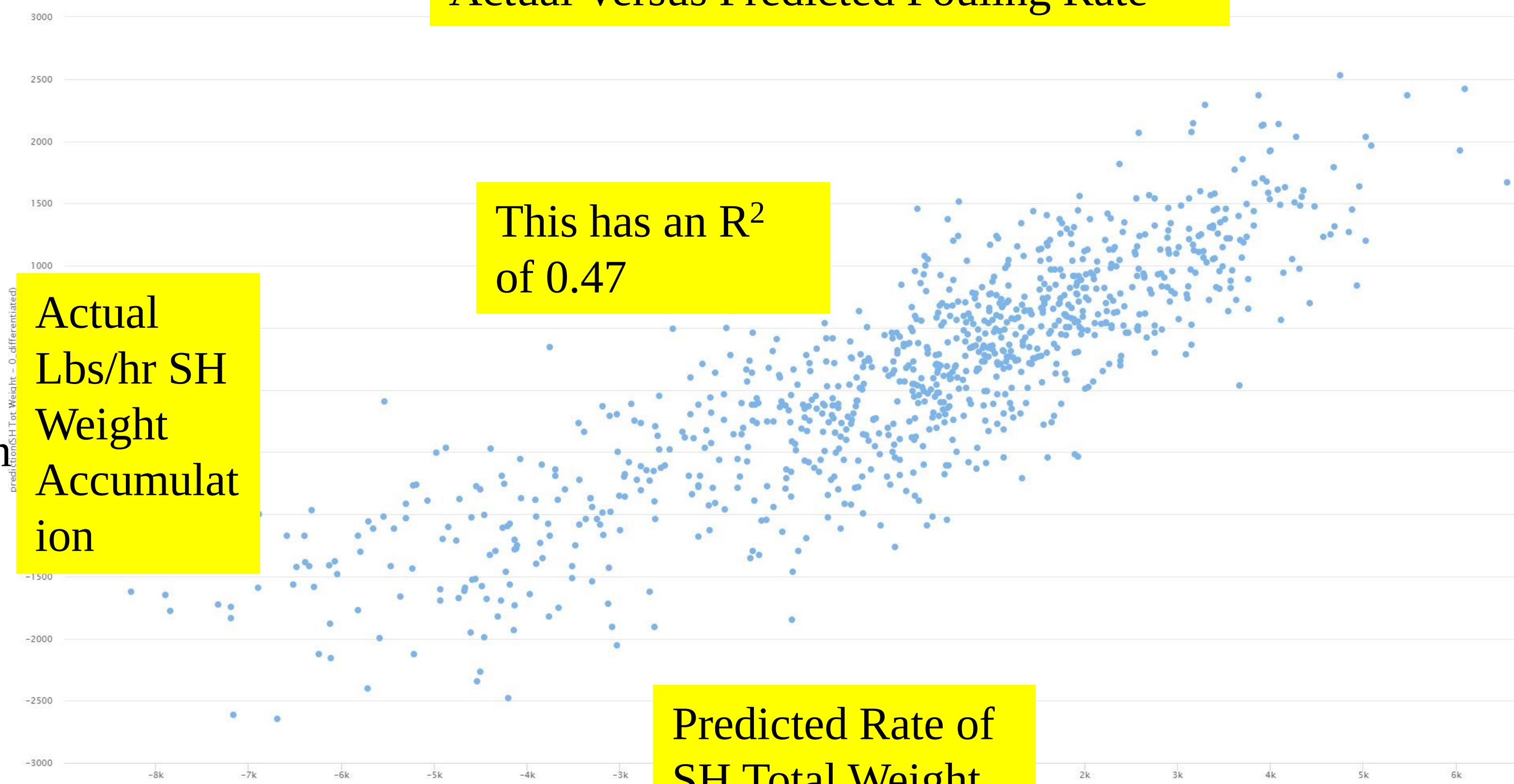
```
Model Metrics Type: Regression
 Description: N/A
 model id: rm-h2o-model-production_model-2906
 frame id: rm-h2o-frame-production_model-2906
 MSE: 3639368.2
 RMSE: 1907.7128
 R^2: 0.472368
 mean residual deviance: 3639368.2
 mean absolute error: 1472.3959
 root mean squared log error: NaN
Variable Importances:
```

The R^2 here is the indication of how well the model was "trained"

Variable	Relative Importance	Scaled Importance	Percentage
TOTAL IK STEAM FLOW E/W 0RFR7074.AVG	24759455744.000000	1.000000	0.16349
5RB SEC AIR HDR FRONT PRESS R_5RPI6905-PV	14672083968.000000	0.592585	0.096884
#5 RB ACTUAL LIQ FLOW 5RFI06203B.PV	12801754112.000000	0.517045	0.084534
RB5 ID FAN speed R_5RSI6494-PV	12563651584.000000	0.507428	0.082962
5RB CORR COMB AT 8% OXYGEN 20M AVG R_5RAI07059_20A-PV	11582675968.000000	0.467808	0.076484
5RB TERT AIR HDR FRONT PRESS R_5RPI6907-PV	11037844480.000000	0.445803	0.072886
5RB TERT AIR DUCT PRESS R_5RPIC6373-PV	9920230400.000000	0.400664	0.065506
Ash Dissolving Tank Temperature R_5RTIC16544-PV	8702906368.000000	0.351498	0.057468
5RB PRI & SEC TOTAL AIR FLOW R_5RFI6338-PV	8571506176.000000	0.346191	0.056600
5RB TERT FRONT (S) AIR FLOW R_5RFIC6917-PV	6824795648.000000	0.275644	0.045066
Percent Tertiary Air	5578608640.000000	0.225312	0.036837
5RB PRI FRONT (S) AIR FLOW R_5RFIC6915-PV	5240470016.000000	0.211655	0.034604
5RB Corrected TRS 12 hr Avg R_5RAI9963_Avg12h	4988245504.000000	0.201468	0.032939
5RB STARTER BNR OIL FLOW 5RFI06459.PV	4502010880.000000	0.181830	0.029728
5RB LIQUOR TEMPERATURE 0RTC6721.PV	3845398784.000000	0.155310	0.025392
Ash rotary feeder current R_5RII06334-PV	3318419968.000000	0.134026	0.021913
5RB LOAD BURNER OIL FLOW 5RFI06460.PV	2076007040.000000	0.083847	0.013708
5RB LOAD BNR OIL PRESS R_5RPIC6458-PV	453404864.000000	0.018312	0.002994
5RB TERT AIR DUCT PRESS R_5RPIC6373-SP	0.000000	0.000000	0.000000

The amount of sootblowing steam used is the most important variable, not that surprising but wait…

Actual Versus Predicted Fouling Rate



This has an R^2 of 0.47

Foul

Clean

Actual Lbs/hr SH Weight Accumulation

Predicted Rate of SH Total Weight Accumulation

- Attribute weights – What is most important in the model?

attribute	weight
TOTAL IK STEAM FLOW E/W 0RFR7074.AVG	2399077...
5RB TERT AIR DUCT PRESS R_5RPIC6373-PV	1465303...
5RB SECONDARY AIR FLOW R_5RFIC6350-PV	1448032...
5RB LIQUOR GUN PRESSURE 0RPR6204.PV	1427048...
#5 RB ACTUAL LIQ FLOW 5RFI06203B.PV	1387350...
Ash Dissolving Tank Temperature R_5RTIC16544-PV	1330563...
RB5 ID FAN speed R_5RSI6494-PV	9584192...
5RB FEEDWATER FLOW 20 MIN AVG R_5R_FEEDWTR_AVG	7740953...
5RB O2 FROM 5RB STACK 5RAR09964.PV	6947406...
5RB Corrected TRS 12 hr Avg R_5RAI9963_Avg12h	6500730...
5RB CORR COMB AT 8% OXYGEN 20M AVG R_5RAI07059_20A-PV	6392997...
5RB PRI AIR HDR FRONT PRESS R_5RPI6903-PV	6293765...
5RB LIQUOR TEMPERATURE 0RTC6721.PV	6242058...
Ash rotary feeder speed R_5RSI6334-PV	5927867...
5RB PRI FRONT (S) AIR FLOW R_5RFIC6915-PV	5443376...
5RB BL DENSITY A 5RDR06207A.PV	5044140...
5RB CORRECTED TRS 5RAR09963A.PV	4151534...
Ash rotary feeder current R_5RII06334-PV	3600620...
5RB QUAT (N) AIR FLOW R_5RFIC18279-PV	3471248...
5RB LOAD BNR OIL PRESS R_5RPIC6458-PV	2384347...
5RB 850# Steam Flow 5RFI6275A.PV	0
5RB TERT AIR DUCT PRESS R_5RPIC6373-SP	0

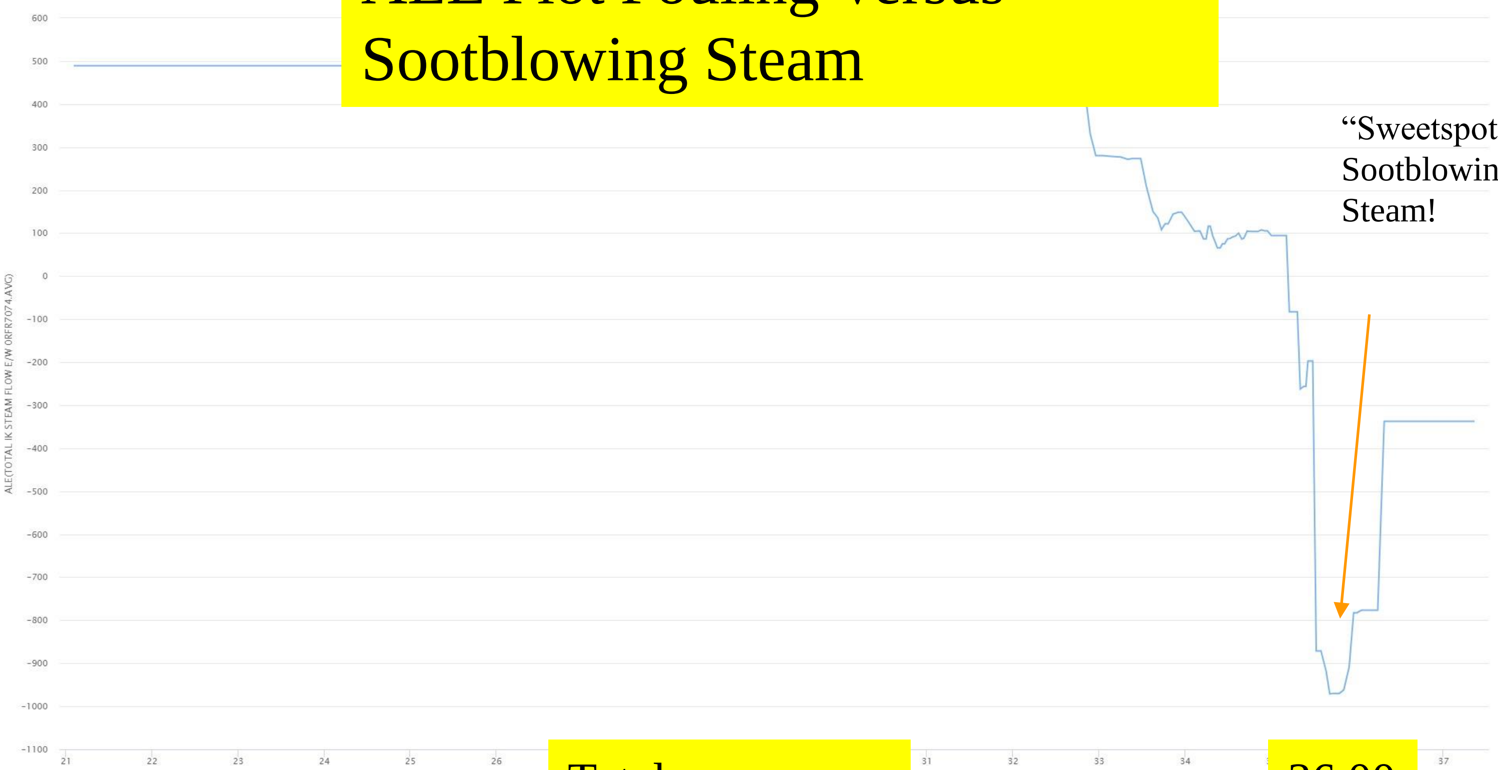
Accumulated Local Effects Plots (ALE)

- One of the biggest challenges of using Gradient Boosted Trees for modeling is interpretation of the results
- You obtain tens to hundreds of trees from the modeling effort, it is not possible to interpret this information from simply studying the trees that are built
  - The most important variables can be identified and this is useful but only to a limited extent
- This is where ALE plots can be used!
- An ALE plot is developed by using a technique that identifies the relationship between a "feature" and the target variable independent of the effects of other features
  - For example it can determine the effect of the "Sootblowing Steam Flow" on the rate of fouling
  - This is an excellent reference on this topic https://christophm.github.io/interpretable-ml-book/ale.html

ALE Plot Fouling Versus Sootblowing Steam

Foul

Clean



"Sweetspot" for Sootblowing Steam!

Total Sootblower Steam Used

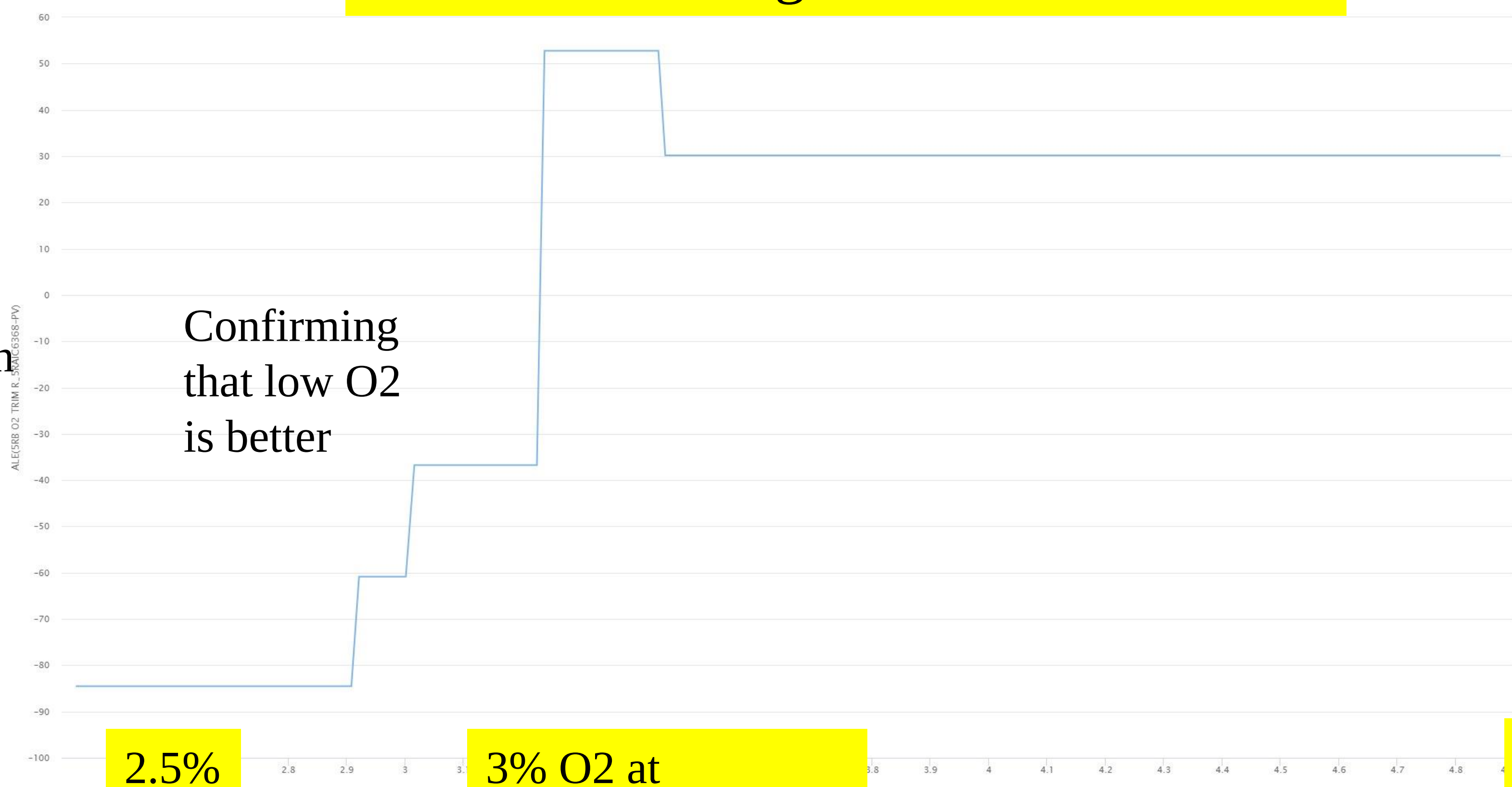
36,000 lbs/hr

SUOMEN SOODAKATTILAYHDISTYS
FINNISH RECOVERY BOILER COMMITTEE

ALE Plot Fouling Versus O2 at Exit

Foul

Clean



Confirming that low O2 is better

2.5%

3% O2 at Economizer Exit

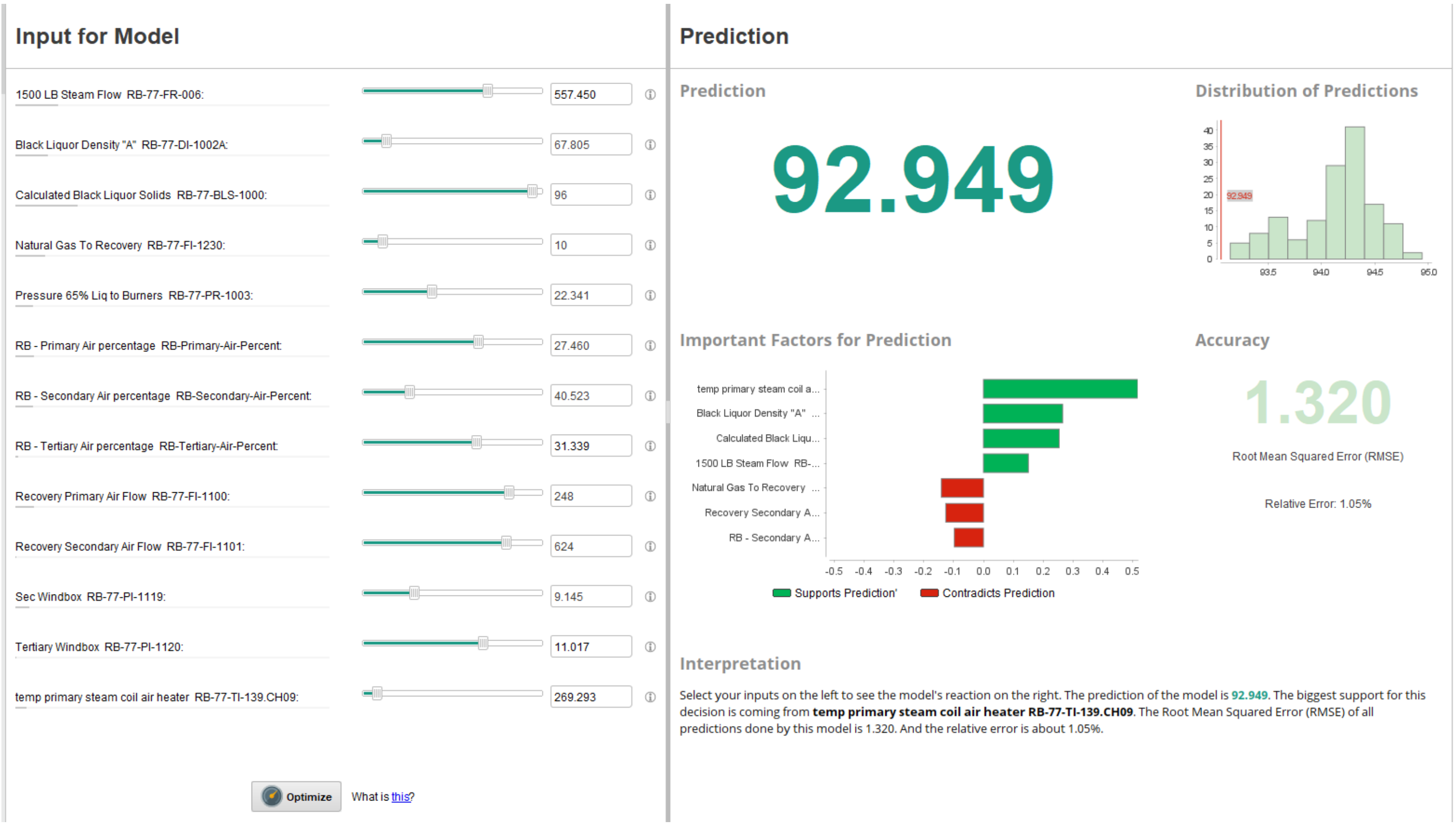
5%

SUOMEN SOODAKATTILAYHDISTYS
FINNISH RECOVERY BOILER COMMITTEE

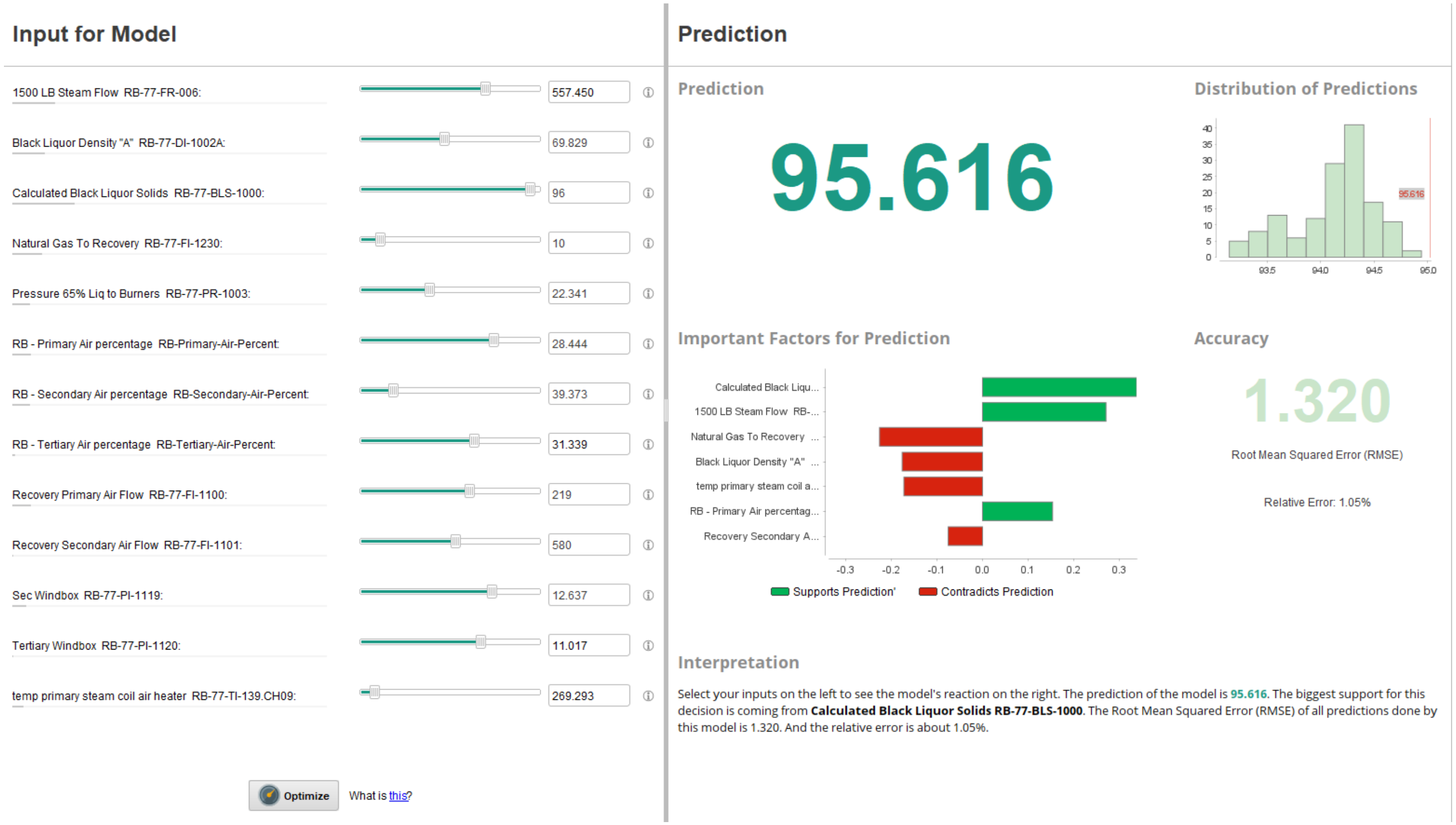
REDUCTION EFFICIENCY OPTIMIZER INTERFACE

Predictive Features for a Prescriptive Model

attribute	weight
Calculated Black Liquor Solids RB-77-BLS-1000	7380.459
temp primary steam coil air heater RB-77-TI-139.CH09	2861.431
Black Liquor Density "A" RB-77-DI-1002A	2822.604
Recovery Secondary Air Flow RB-77-FI-1101	1683.446
Sec Windbox RB-77-PI-1119	1588.928
Natural Gas To Recovery RB-77-FI-1230	1404.008
Tertiary Windbox RB-77-PI-1120	1289.583
1500 LB Steam Flow RB-77-FR-006	1238.175
Recovery Primary Air Flow RB-77-FI-1100	1023.692
RB - Secondary Air percentage RB-Secondary-Air-Perce.	937.234
Pressure 65% Liq to Burners RB-77-PR-1003	933.642
RB - Tertiary Air percentage RB-Tertiary-Air-Percent	867.574
RB - Primary Air percentage RB-Primary-Air-Percent	753.221


Input for Model
1500 LB Steam Flow RB-77-FR-006: 557.450
Black Liquor Density "A" RB-77-DI-1002A: 67.805
Calculated Black Liquor Solids RB-77-BLS-1000: 96
Natural Gas To Recovery RB-77-FI-1230: 10
Pressure 65% Liq to Burners RB-77-PR-1003: 22.341
RB - Primary Air percentage RB-Primary-Air-Percent: 27.460
RB - Secondary Air percentage RB-Secondary-Air-Percent: 40.523
RB - Tertiary Air percentage RB-Tertiary-Air-Percent: 31.339
Recovery Primary Air Flow RB-77-FI-1100: 248
Recovery Secondary Air Flow RB-77-FI-1101: 624
Sec Windbox RB-77-PI-1119: 9.145
Tertiary Windbox RB-77-PI-1120: 11.017
temp primary steam coil air heater RB-77-TI-139.CH09: 269.293
Optimize
What is this?
Prediction
Prediction
92.949
Distribution of Predictions
92.949
Important Factors for Prediction
temp primary steam coil a...
Black Liquor Density "A" ...
Calculated Black Liqu...
1500 LB Steam Flow RB-...
Natural Gas To Recovery ...
Recovery Secondary A...
RB - Secondary A...
-0.5 -0.4 -0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4 0.5
Supports Prediction'
Contradicts Prediction
Accuracy
1.320
Root Mean Squared Error (RMSE)
Relative Error: 1.05%
Interpretation
Select your inputs on the left to see the model's reaction on the right. The prediction of the model is 92.949. The biggest support for this decision is coming from temp primary steam coil air heater RB-77-TI-139.CH09. The Root Mean Squared Error (RMSE) of all predictions done by this model is 1.320. And the relative error is about 1.05%.

SUOMEN SOODAKATTILAYHDISTYS
FINNISH RECOVERY BOILER COMMITTEE



Issues With Data That Must be Considered

Autocorrelation

- Prior target values show high correlation, especially with frequent test data and downstream of larger storage tanks. Be cautious during modeling to ensure accuracy and select appropriate model types.

Multi-collinearity

- Many of the features you choose are correlated with other features
- Without attention to this models can be highly sensitive to very small changes in highly correlated features
- Careful removal of correlated features is required

Missing values

- When to drop data that has missing values must be considered
- How to deal with missing values in your target variable (interpolation?)
- How to address missing values in your features (averaging?)

Lagged

- Upstream variables are often DYNAMICALLY lagged versus your target variable
- What is the lag and how do you handle this if you were to deploy a model built from this data

How Do You Identify a Feature Set For A Model?

- In Type 1 Models (Forecast Type) – you can select from a wide-variety of features
  - Co-variant with your target (not causal)
  - Causal variables – known control action on target variable
  - Possibly causal variables – suspected influence on target variable
  - Prior values of your target variable
- In Type 2 Models (Prescriptive) – Need to be more careful and can choose only causal variables - Excellent domain knowledge is needed

There's a trade-off between model accuracy and usefulness. Models with fewer "critical" features are more accurate and stable but may miss important relationships. Various methods like "Backwards Elimination," "Correlation," and "Added Information" help determine the best feature set. It's wise to try different methods.

A Brief History of the use of Machine Learning

- Many of the techniques used have been invented within the past 20 years
- Academic to industrial use 20→10 years
- Practical on laptop computers and ability to deploy models on industrial systems 10→5 years
- Training of data scientists 10→5 years
- Development of applicable examples and deployment in operating environments 5→0 years
- Developing experience with application, integrating into operating practice 5→ Future