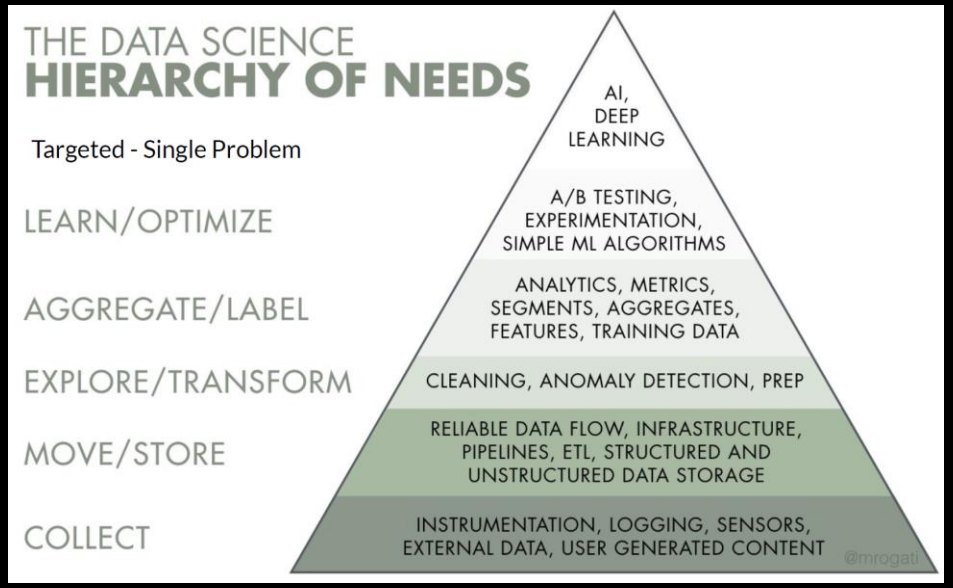
TRIPLE POINT

No-code 6 Step Machine Learning

Load Data → Feature Engineering → View Data → Send Data → Train Model → Test Model

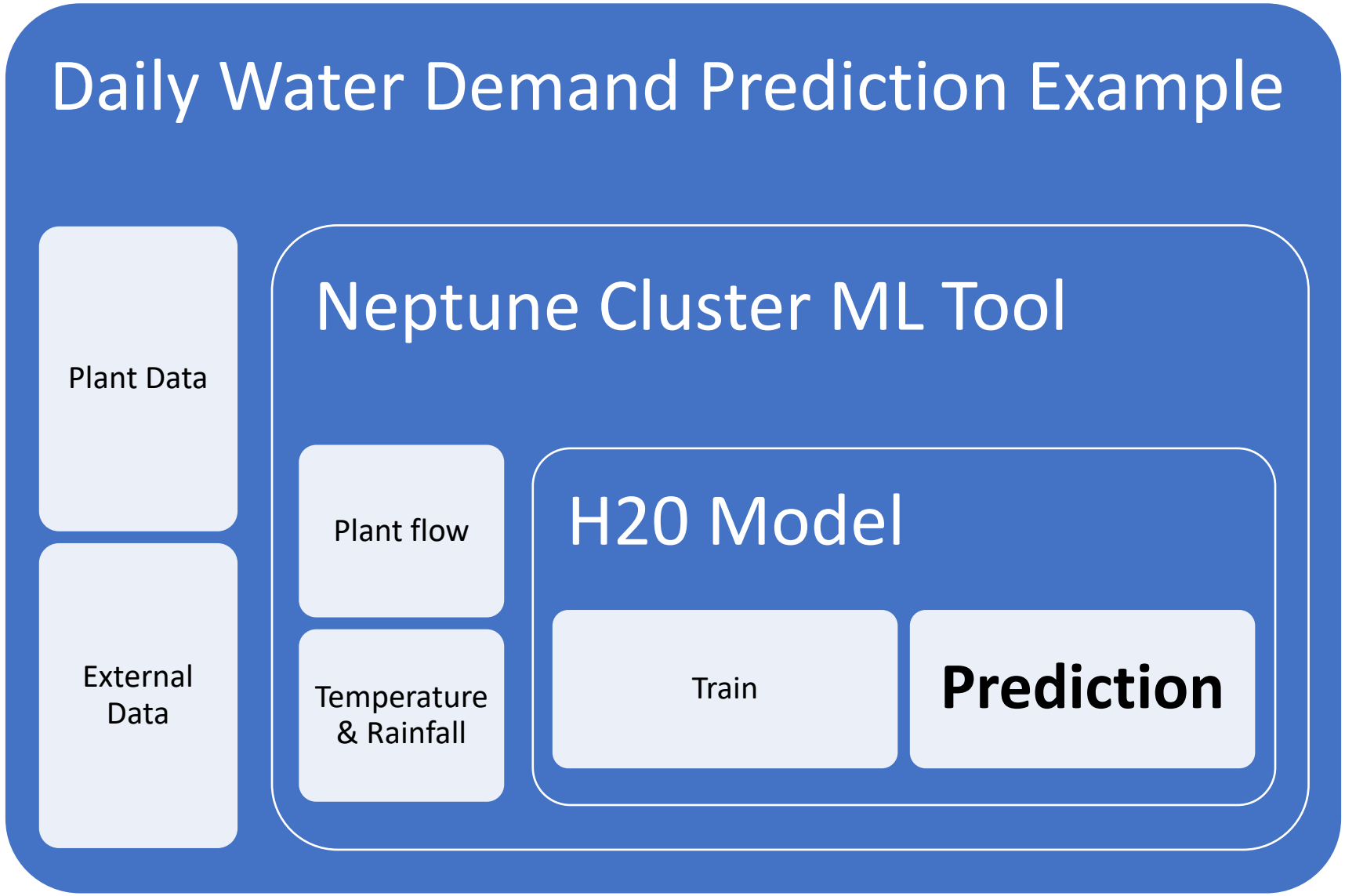
NEPTUNE CLUSTER ML



When scaling, group all elements to be scaled. Scale as needed. Use the “Increase Font Size,” “Decrease Font Size” buttons or manually change the font size for the editable text. Delete unwanted elements. Change the line weight of the circles or connecting bars as needed.

TRIPLE POINT

No-code 6 Step Machine Learning



Step 1: Connect to the Trident Historizer and choose your time period and tags

In this example, a start date of January 1st 2013 is selected with a time frame of eight years with the timeseries data average to one day periods. The tags selected are water flow out of the treatment plant along with weather data in the area (temperature and rainfall)

1. Trident Historizer

URL

192.168.1.140:8080

2. Query Selection

Choose Start Date

2013/01/01

timevalue

8

timeunits

- days
- weeks
- months
- years

pre-aggregators

avg

value

1

units

- minutes
- hours
- days
- weeks
- months
- years

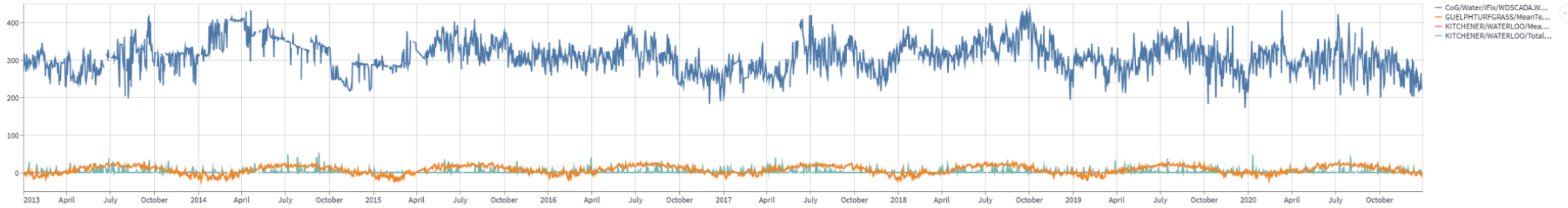
Neptune Cluster Machine Learning

1. Data loading

Select model inputs

CoG/Water/iFix/WD... | KITCHENER/WATER... | KITCHENER/WATER... | GUELPHTURFGRAS...

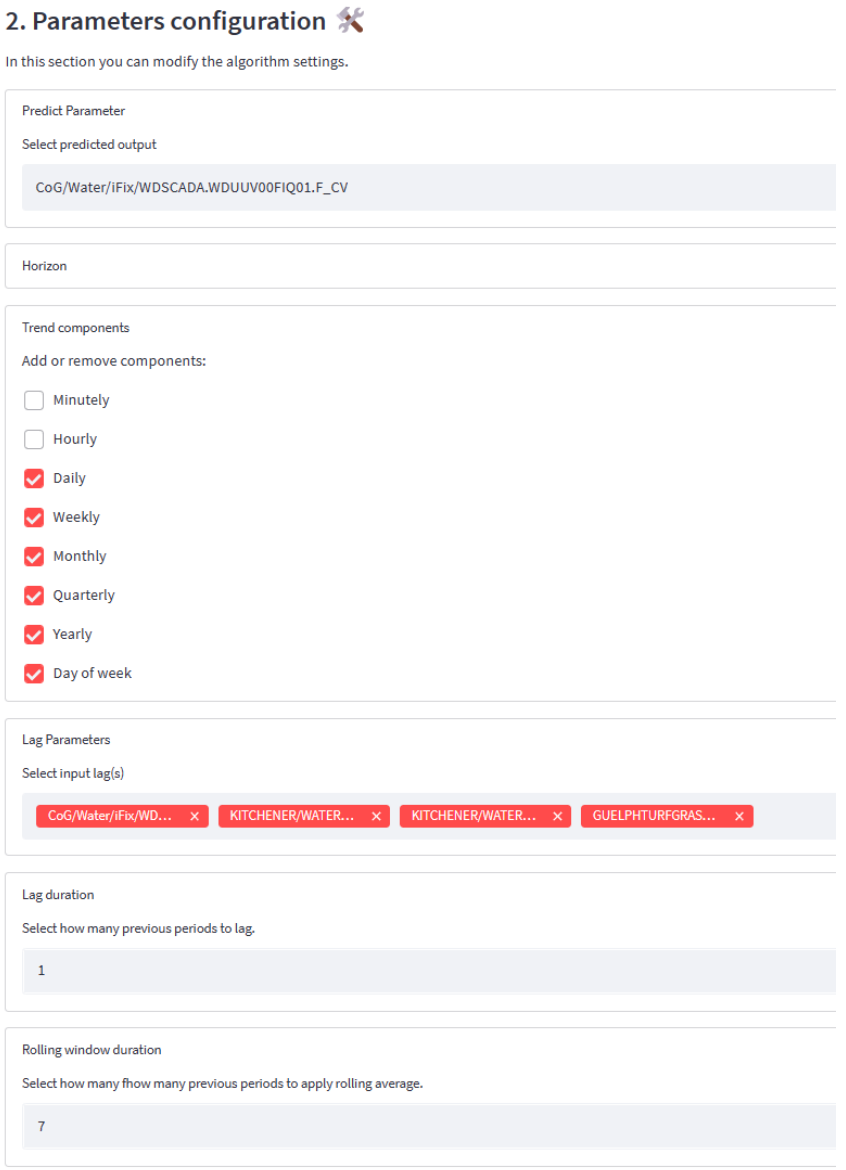
	CoG/Water/iFix/WDSCADA.WDUUV00FIQ01.F_CV	KITCHENER/WATERLOO/MeanTemp	KITCHENER/WATERLOO/TotalPrecip1	GUELPHTURFGRASS/MeanTemp
count	2,753.0000	2,754.0000	2,760.0000	2,834.0000
mean	308.1125	7.2143	2.1233	6.8766
std	41.2590	10.7812	5.3699	10.7329
min	171.3367	-24.4000	0.0000	-23.6000
25%	282.2329	-1.0000	0.0000	-1.3000
50%	307.3308	7.5000	0.0000	7.2000
75%	333.6922	16.7000	1.2000	16.2000
max	432.1210	27.3000	53.0000	27.3000



TRIPLE POINT

Step 2: Setup your parameter configuration
-Choose your predicted parameter
-Choose your trend components (feature engineering)
-Select your lag parameters and durations

In this example, the timeseries water flow out of the treatment plant is chosen as the predicted parameter. All trend components are selected except for minutely and hourly since they provide no value into a daily prediction. All tag components are selected as lag parameters. The lag period is set to 1 day and the rolling average is set to 7 days. These parameters can be adjusted to improved performance.



TRIPLE POINT

Step 3: Review your dataset and send to model

3. View Data

	CoG/Water/iFix/WDSCAD...	KITCHENER/WATERLOO/...	KITCHENER/WATERLOO/T...	GUELPHTURFGRASS/Mea...	week	quater	year	month	day	lag_CoG/Water/iFix/WDS...	rolling_mean_CoG/Water...	lag_KITCHENER/WATERL...	rolling_mean_KITCHENE...	lag_KITCHENER/WATE
count	2,753.0000	2,754.0000	2,760.0000	2,834.0000	2,834.0000	2,834.0000	2,834.0000	2,834.0000	2,834.0000	2,752.0000	2,412.0000	2,753.0000	2,419.0000	2,759
mean	308.1125	7.2143	2.1233	6.8766	26.7029	2.5124	2,016.5042	6.5480	15.6486	308.1189	306.8320	7.2194	6.9811	2
std	41.2590	10.7812	5.3699	10.7329	15.0473	1.1169	2.3095	3.4445	8.7918	41.2651	34.3896	10.7799	10.0718	5
min	171.3367	-24.4000	0.0000	-23.6000	1.0000	1.0000	2,013.0000	1.0000	1.0000	171.3367	216.6863	-24.4000	-19.6571	0
25%	282.2329	-1.0000	0.0000	-1.3000	14.0000	2.0000	2,014.0000	4.0000	8.0000	282.2280	283.4419	-1.0000	-1.2071	0
50%	307.3308	7.5000	0.0000	7.2000	27.0000	3.0000	2,016.0000	7.0000	16.0000	307.3454	306.0483	7.5000	6.9143	0
75%	333.6922	16.7000	1.2000	16.2000	40.0000	4.0000	2,019.0000	10.0000	23.0000	333.7125	326.3948	16.7000	16.2786	1
max	432.1210	27.3000	53.0000	27.3000	53.0000	4.0000	2,020.0000	12.0000	31.0000	432.1210	408.3561	27.3000	24.9143	53

Exclude Parameters +

4. Send Data

Send Data

In this example, the data count is 2753 (7,928,640 rolled up flow points!) . 8 years x 365 days = 2920. So, we are short a few random days. That's OK. Data is sent to the model for training. Training is setup to test against 25 different models. Each model will run for a max time of 120 seconds and the whole runtime is set to 300 seconds.

Step 4: Train your model

5. Train Model

AutoML Parameters −

Select how many models to generate.

25 − +

Select maximum runtime in seconds.

300 − +

Select maximum runtime per model in seconds.

120 − +

Run

Done!

TRIPLE POINT

Step 5: Review the performance of you different models

The model leader board shows the performance of each model that ran ranked from top performer to worst performer. These a characterized by root mean square error, mean square error, mean absolute error, root mean squared logarithmic error and mean residual deviance.

5. Model Leader Board

	model_id	mean_residual_deviance	rmse	mse	mae	rmsle
0	StackedEnsemble_BestOfFamily_7_AutoML_2_20221215_144539	491.7186	22.1747	491.7186	16.3801	0.0749
1	StackedEnsemble_BestOfFamily_4_AutoML_2_20221215_144539	491.8218	22.1771	491.8218	16.3887	0.0749
2	StackedEnsemble_AllModels_3_AutoML_2_20221215_144539	494.3311	22.2336	494.3311	16.4792	0.0753
3	StackedEnsemble_AllModels_1_AutoML_2_20221215_144539	495.4816	22.2594	495.4816	16.5263	0.0752
4	StackedEnsemble_BestOfFamily_2_AutoML_2_20221215_144539	495.7036	22.2644	495.7036	16.4678	0.0751
5	StackedEnsemble_BestOfFamily_3_AutoML_2_20221215_144539	496.2389	22.2764	496.2389	16.5219	0.0753
6	StackedEnsemble_AllModels_4_AutoML_2_20221215_144539	496.8457	22.2900	496.8457	16.4582	0.0754
7	StackedEnsemble_AllModels_7_AutoML_2_20221215_144539	497.0415	22.2944	497.0415	16.4560	0.0754
8	StackedEnsemble_AllModels_2_AutoML_2_20221215_144539	497.4873	22.3044	497.4873	16.5527	0.0754
9	GBM_grid_1_AutoML_2_20221215_144539_model_2	499.2520	22.3439	499.2520	16.4477	0.0756

Step 6: Test your different models

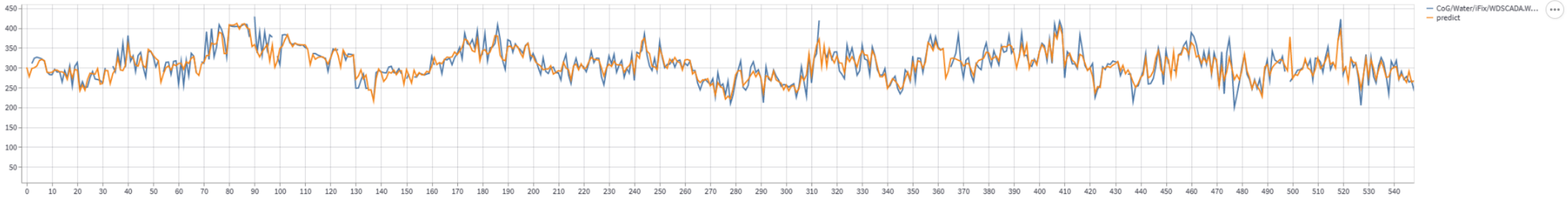
Using the drop-down selector, the different models can be tested on the various portions of the dataset. The dataset was divided into 80% model generation and 20% model testing. The test dataset is shown first, then the 20% model testing and finally the full data set last. The blue line is the actual water demand and the orange line is the model generated predicted demand.

6. Test Model - Make a Prediction

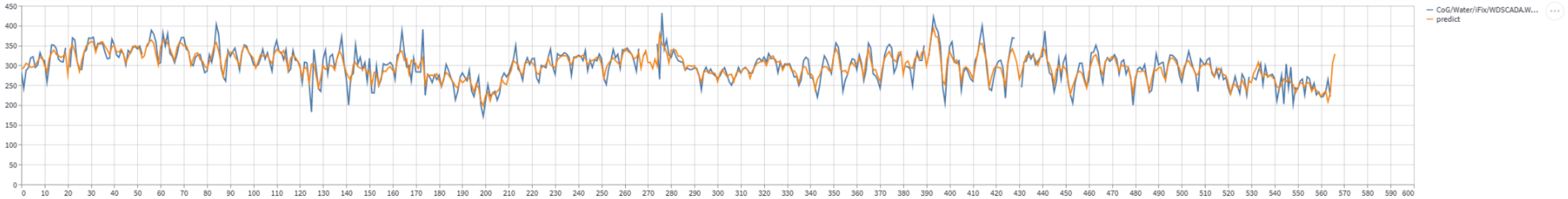
Select model

StackedEnsemble_BestOfFamily_7_AutoML_2_20221215_144539

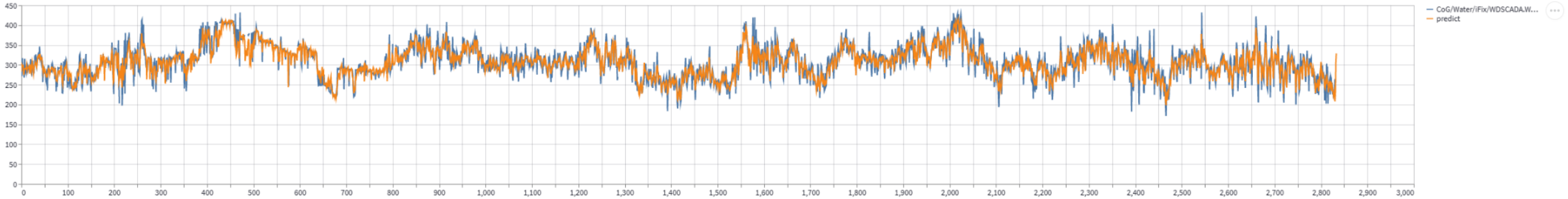
Test Dataset



Last 20% of Dataset



Full Dataset



TRIPLE POINT

NEPTUNE CLUSTER ML

You're Now Ready to Transition to Production!