#### Video Place Holder



## Deep Dive: How to use Splunk to protect the environment

Charlie Studdard

Machine Data Analytics | Honda

Manufacturing of Alabama



#### **Forward-Looking Statements**

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

#### My Honda Story

My Experience at Honda Manufacturing of Alabama

- Core Values
  - Challenging Spirit
    - I feel I can make a difference everyday by providing insights to the company that no one has ever provided before.
  - Dreams
    - I am constantly trying to come up with ideas to help improve not only our products but also how to help make better business by building out a connected and smart plant.

Dates	Title	Role	Description
02/2007	Contractor	Help Desk	Took and solved calls from the business side and manufacturing side environments
08/2008	Contractor	PC Services	Managed IT assets for over 4500 Associates and Contractors
03/2011	Associate Technical Specialist	Shift Lead	Lead for Help Desk Team
04/2014	Engineer Specialist	Plant Floor Systems	Developed and supported our assembly line control system
04/2016	Engineer Specialist	DBA	Administered the manufacturing side DB2 databases. Also, built and administered our Splunk environment. Worked with multiple departments to bring in different types of data
04/2019	Staff Engineer	Data Analytics	Use machine data to help proactively take steps to reduce downtime and failures

#### **Shout Out**

#### I could not have done this without

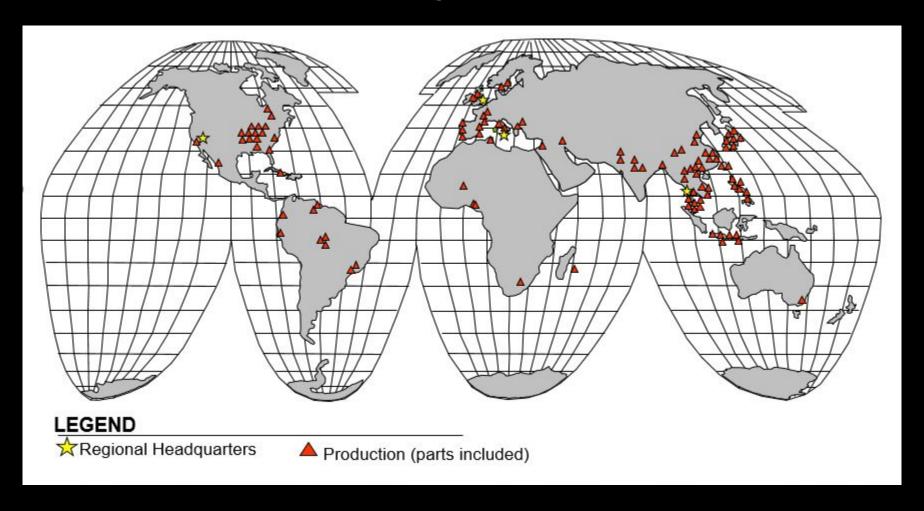
- Geoff Duncan- Honda Dept Manager
- Bobby Rogers- Honda Senior IT Architect
- Chris McLain- Honda LTM
- Matt McLaren- Splunk Account Manager
- Wyatt Banks- Splunk Sales Engineer
- Corey Whitney- Honda Engineer
- Matt Thielker- Honda Engineer
- Gyanendra Rana- Splunk Machine Learning Advisory Lead
- Zidong Yang- Splunk Machine Learning Data Scientist
- Rob Gustafson- Splunk Sales Engineer

## "The day I stop dreaming is the day I die"

Soichiro Honda

#### **Honda Global**

150 Manufacturing Facilities Worldwide



### Quick Facts About Honda Manufacturing of Alabama







• Associates 4,500

• Plant size 4.2 million square feet

• Site size 1,350 acres

Capital Investment \$2.6 billion

Models

 Odyssey, Pilot, Ridgeline,
 Passport and V-6 engines

Production Capacity 340,000 vehicles & engines

2 and 3 Shift Operations



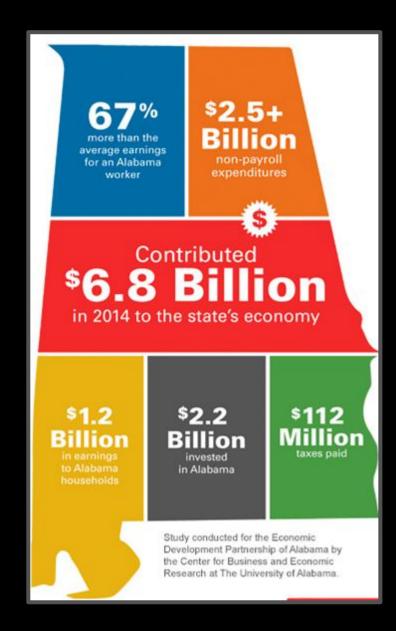


#### What is a manufacturing plant?

- At HMA we make the vehicle:
  - Take metals like iron and aluminum, melt them down and cast them into engine blocks and heads.
  - Machine metal to create crankshafts and conrods
  - Assemble these parts to create engines.
  - Stamp metal coils to create the side panels, doors, hoods, tailgates, etc.
  - Weld these parts together to create bodies
  - Paint these welded bodies
  - Then take thousands of other parts to finish assembling each vehicle
  - Finally test and inspect every vehicle before shipping it to a dealership

## Economic Impact of Honda Manufacturing of Alabama

- In 2014 HMA contributed \$6.8 Billion to Alabama's economy.
- \$2.2 Billion invested into various industries and programs in Alabama
- ▶ \$1.2 Billion to earnings to Alabama Households.



#### Splunk Overview

#### A brief history of our Splunk uses

- Early 2017- Began using Splunk for application log analysis and monitoring
- .conf17- Realized it could be so much more
- Early 2018- Began Splunking VFD(variable frequency drive) data, giving equipment associates their first real time and historical look at equipment data.
- Late 2018- Approached by the Paint Department to record and backup data for the RTO(regenerative thermal oven), which is used to oxidize pollutants such as VOCs.
- Early 2019- Start engagement with Machine Learning Advisory
- Today- Implemented new dashboards with power, water, natural gas, and compressed air for the plant.

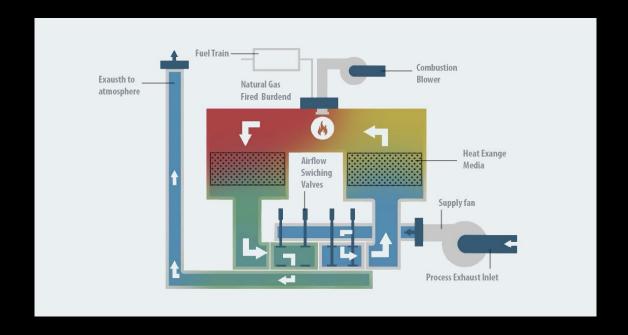


#### **Project Goal**

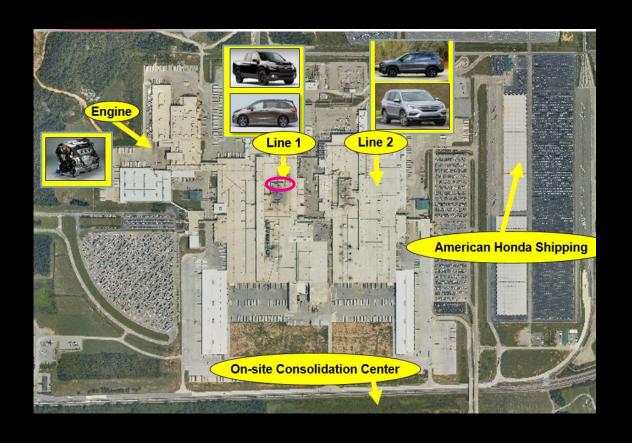
- Can we use Splunk's Machine Learning Toolkit to predict issues with the RTO?
  - If so how far out can we push the prediction?

#### What is a RTO?

- More properly known as a Regenerative Thermal Oxidizer, it takes exhaust fumes and burns off pollutants like VOCs.
- A high pressure fan forces exhaust through ceramic heat exchange media to preheat the exhaust. Then it reaches the combustion chambers and is held until the temperature reaches oxidation for the pollutants (1525° F).
- After the combustion chamber the exhaust is moved through another ceramic heat exchange media to help cool off the air and capture heat to be reused.
- If the RTO does not reach this temperature, it could release VOCs in the environment.



#### **Plant Map**





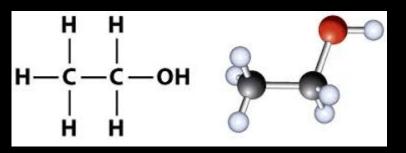
#### What is a VOC?

#### Volatile Organic Compound

#### **EPA Definition**

- Volatile organic compounds, or VOCs are organic chemical compounds whose composition makes it possible for them to evaporate under normal indoor atmospheric conditions of temperature and pressure.
   Some of which are known carcinogens.
- Examples
  - Propane
  - Butane
  - Formaldehyde
  - Ethanol

#### **Ethanol Structure**

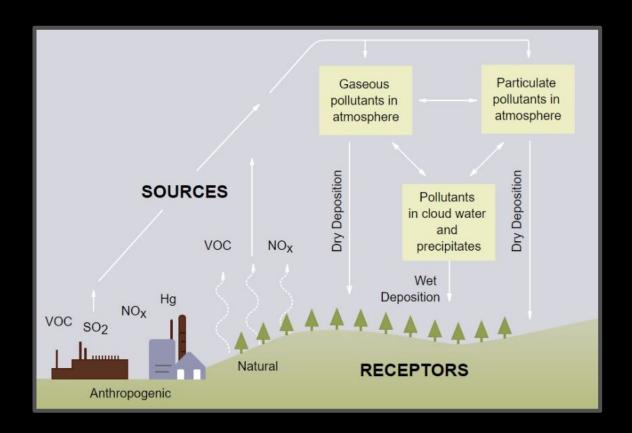


### How do VOCs impact the environment?

Both indoor and outdoor

Preathing VOCs can irritate the eyes, nose and throat, can cause difficulty breathing and nausea, and can damage the central nervous system as well as other organs. Some VOCs can cause cancer. Not all VOCs have all these health effects, though many have several.

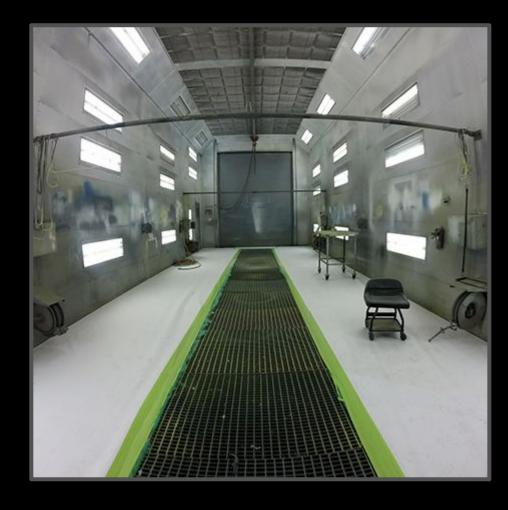
source-https://www.lung.org/our-initiatives/healthy-air/indoor/indoor-air-pollutants/volatile-organic-compounds.html



## How do VOCs impact the environment?

**Indoor Impact** 

We paint the vehicles inside paint booths. The booths have a ventilation system that takes the paint fumes and moves them to the RTO.



#### How does it impact the business?



If the RTO goes down, no vehicles can be painted.
Which could cause workers to be sent home early.



The plant is JIT (just in time) delivery, and downtime impacts the supply chain in adverse ways.



Not just at the plant, all the suppliers that provide parts might have to stop their production to allow supply chain stoppages

to resolve.

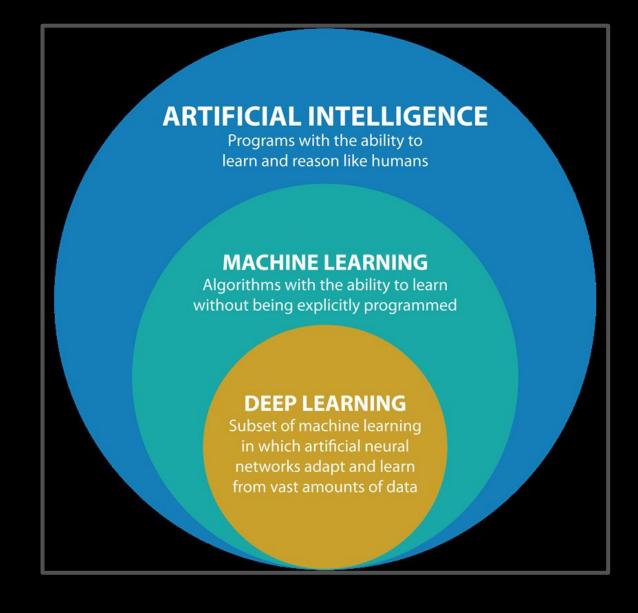
# "Instead of being afraid of the challenge and failure, be afraid of avoiding the challenge and doing nothing"

Soichiro Honda

### What is Machine Learning?

And how does it relate to Artificial Intelligence?

- Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed.
- Machine Learning (ML) is a subset/precursor to AI. ML informs or helps realize AI.
- Side Note- Deep Learning is a specific discipline of ML that has seen a surge of breakthroughs in the past decade.



#### How did I become involved with ML?

#### What got me interested in ML?

- Started to see articles about Google's AlphaGo Zero and OpenAl's success in beating Human players in games.
- Found youtube channels about using OpenAI gym to work on your own ML projects. Sethblings MarI/O and Mario Kart videos, Lucas Thomas' OpenAI videos, and Siraj Raval's channel.
- Kai-Fu Lee's book- Al Superpowers: China, Silicon Valley, and the New World Order
- This started my own personal project to get train a model to complete Super Mario Bros Level 1-1.

#### Where did I learn about ML?

- Coursera's Machine Learning course by Andrew Ng.
- Working with Splunk's Machine Learning Advisory Team.
- Book: Deep Learning with Python by Francois Chollet.
- Websites: Hackernoon.com, TowardsDataScience.com, kdnuggets.com, arXiv.org

#### Splunk Machine Learning Advisory Program



- Get help from the Splunk Data Scientists to solve your business use case with Machine Learning Toolkit
- Complimentary support with your Enterprise or Cloud license
- Early access to new Machine Learning features
- Results in opportunity to tell your success story with Splunk
- Contact <u>mlprogram@splunk.com</u> for more information

#### Splunk's Machine Learning Advisory Program

#### Splunk's Team

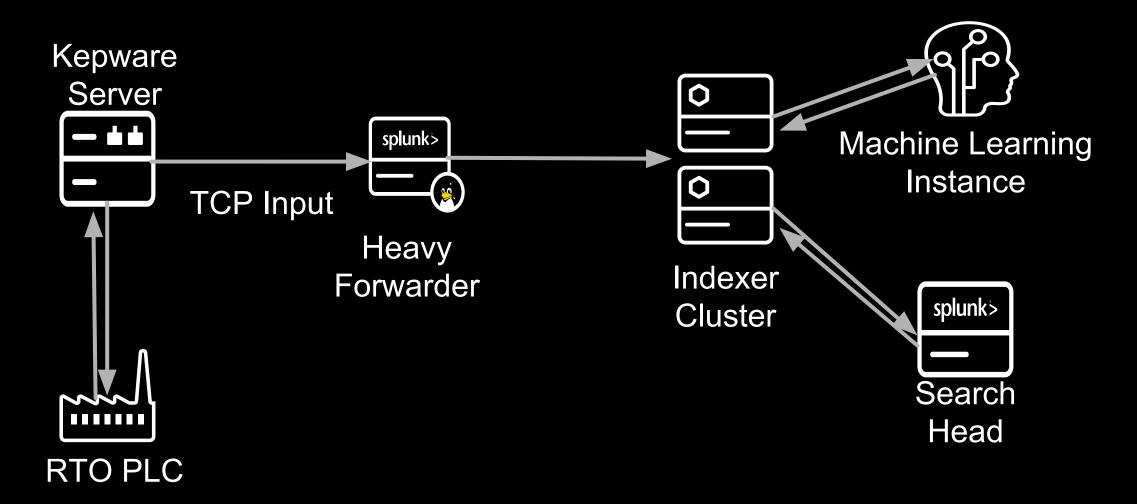
- Around summer 2018 we started to discuss the data we were collecting with our Account Manager, Matt McLaren, and Sales Engineer, Wyatt Banks, and they mentioned we might be able to work with Splunk's Machine Learning Advisory Program.
- Then early this year we had our first meeting with Gyanendra and his team. After going over the machine data we were collecting and the goal of the project we decided we should try to tackle predictive failures with the RTO.

#### Why RTO?

- Out of all the equipment data we collect, it offers the most diverse data.
  - Motor Side Vibrations
  - Fan Side Vibrations
  - Exhaust Fan Motor A Phase Temp
  - Exhaust Fan Motor B Phase Temp
  - Exhaust Fan Motor C Phase Temp
  - Exhaust Fan Speed
  - Exhaust Fan Side Temp
  - Exhaust Motor Side Temp
  - Chamber A Temp
  - Chamber B Temp



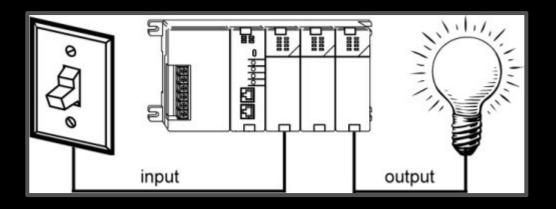
#### How we get the data



#### What is a PLC?

#### Programmable logic controller

PLC) or programmable controller (PLC) or programmable controller is an industrial digital computer which has been ruggedized and adapted for the control of manufacturing processes, such as assembly lines, or robotic devices, or any activity that requires high reliability control and ease of programming and process fault diagnosis.





#### What does the data look like?

>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Motor_Side_Vibrations" Value="76" Quality="good"  host = hmaapp251.am.mds.honda.com   source = kepware   sourcetype = plc:PA1
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Fan_Side_Vibrations" Value="31" Quality="good"  host = hmaapp251.am.mds.honda.com   source = kepware   sourcetype = plc:PA1
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Exhaust_Fan_Speed" Value="48" Quality="good"  host = hmaapp251.am.mds.honda.com
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Exhaust_Fan_Side_Temp" Value="141" Quality="good" host = hmaapp251.am.mds.honda.com   source = kepware   sourcetype = plc:PA1
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Exhaust_Fan_Motor_Temp_C_Phase" Value="178" Quality="good" host = hmaapp251.am.mds.honda.com   source = kepware   sourcetype = plc:PA1
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Exhaust_Fan_Motor_Temp_B_Phase" Value="177" Quality="good" host = hmaapp251.am.mds.honda.com   source = kepware   sourcetype = plc:PA1
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Exhaust_Fan_Motor_Temp_A_Phase" Value="177" Quality="good" host = hmaapp251.am.mds.honda.com   source = kepware   sourcetype = plc:PA1
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Exhaust_Fan_Motor_Side_Temp" Value="145" Quality="good" host = hmaapp251.am.mds.honda.com
>	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Chamber_B_Temp" Value="1561" Quality="good"  host = hmaapp251.am.mds.honda.com
^	7/30/19 11:09:05.998 AM	2019-07-30 16:09:05.998 +0000 Tag="PA1_RTO.RTO.Chamber_A_Temp" Value="1559" Quality="good"  host = hmaapp251.am.mds.honda.com   source = kepware   sourcetype = plc:PA1

#### **Challenges This Data Presents**



Every tag (aka feature) is a separate event.



Weekend data was negatively impacting models because there was no work being done.



Not every tag was collected every scan.

#### **Challenge #1**

How to get all the features into one event

#### Challenge #2

#### Get rid of weekend data

```
index=plc_data source=kepware sourcetype="plc:pa1" Tag="PA1_RTO.RTO.Motor*" OR Tag="PA1_RTO.RTO.Chamber*" OR Tag="PA1_RTO.RTO.Fan*" OR Tag="PA1_RTO.RTO.Exhaust*" Value<65000

| eval day_of_week = strftime(_time, "%A")

| where NOT (day_of_week="Saturday" OR day_of_week="Sunday")

| fields + motor_side_vibrations fan_side_vibrations chamber_b_temp chamber_a_temp exhaust_fan_motor_temp_a_phase exhaust_fan_motor_temp_b_phase exhaust_fan_motor_temp_c_phase
| exhaust_fan_speed exhaust_fan_side_temp exhaust_motor_side_temp

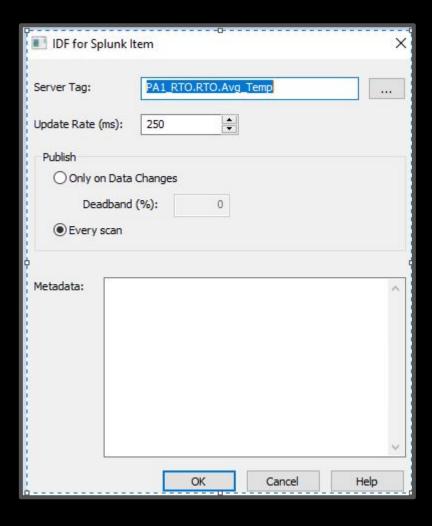
| filldown
| where isnotnull(exhaust_fan_motor_temp_b_phase) ANO isnotnull(chamber_b_temp) ANO isnotnull(fan_side_vibrations) ANO isnotnull(chamber_a_temp)

| ANO isnotnull(exhaust_fan_motor_temp_c_phase) ANO isnotnull(motor_side_vibrations) ANO isnotnull(exhaust_motor_side_temp)
```

#### Challenge #3

Not all tags/features were collected every scan

This was a Kepware setup issue. We had Kepware set to only send data to Splunk on data change. That was changed to send data every cycle.



#### Results

After cleaning up the data what were the results of predicting Chamber A temp use random forest regressor?

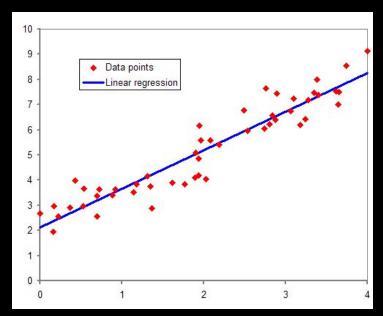


#### What is a regressor?

It is an approach to model the relationship between a dependent variable and independent variable

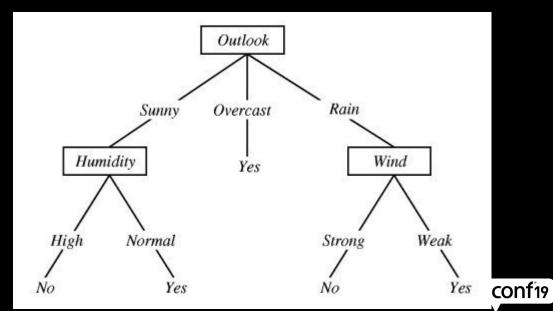
#### **Simple Linear Regression**

Predict the target variable Y with input X. e.g. Housing Prices, take the square footage(X) and predict the price(Y) (known as features). Linear Regression is a supervised learning problem because you need a data set with real data. Y = a + bX



#### **Random Forest Regressor**

PRFR is basically a decision tree that is bootstrapped/split on different features or groups of features. Why did RFR work so well? Because the strength of RFR lies in its versatility with high dimensionality and non-linear data.



#### Why machine learning?

We already have an alerting service/platform.

- Ask the person/team that setup the current/old alerting solution how much fun it was and how they figured out the thresholds to alert on.
  - When we were searching for an alerting system replacement, the number one requirement was dynamic-thresholding. My team manager at the time was the person who setup our old monitoring system.
- Does your current system take in seemingly unrelated data points like dew points, humidity, peak torque voltages, and tell you how closely coupled these points are with your target variable?
- Can your old/current monitoring solution give you the ability to be proactive and not reactive?

# "We only have one future, and it will be made of our dreams, if we have the courage to challenge convention"

Soichiro Honda

#### How to use the data

With such great results from the random forest regressor, what else can we do?



#2



Predict chamber temperatures 24 hours in advanced.

Take important features and look for anomalies to try to predict equipment failure.

Give the equipment engineers a visual to guide them on the status of the RTO.

#### **Use Case #1**

#### How does the data hold up to 8/24/48 hour predictions?

- Working with the Splunk Machine Learning Advisory team, they suggested to try a relative time shift forward in the data. We achieved 0.77 r<sup>2</sup> and 4.94 RMSE for a 24 hour relative shift. Anything beyond 24 hours was well below 0.65 r<sup>2</sup>.
- We also used StateSpace to try to forecast the chamber temperatures. Since StateSpace requires continuous data, our previous way of excluding weekend data had to be augmented with the imputer algorithm to transform the data so StateSpace could be used.
- We also setup an alert around the residual. It might be useful to know when the fit job residual jumps higher.

```
index=plc_data source=kepware sourcetype="plc:pa1" Tag="PA1_RTO.RTO.Motor*" OR Tag="PA1_RTO.RTO.Chamber*" OR Tag="PA1_RTO.RTO.Fan*" OR Tag="PA1_RTO.RTO.Exhaust*" Value<65000
I where NOT (day of week="Saturday" OR day of week="Sunday")
    | fields + motor_side_vibrations fan_side_vibrations chamber_b_temp chamber_a_temp exhaust_fan_motor_temp_a_phase exhaust_fan_motor_temp_b_phase exhaust_fa
                   exhaust fan sneed exhaust fan side temm exhaust motor side temm
   where isnotnull(exhaust_fan_motor_temp_b_phase) AND isnotnull(chamber_b_temp) AND isnotnull(fan_side_vibrations) AND isnotnull(chamber_a_temp) AND isnotnull(exhaust_fan_motor_temp_b_phase) AND isnotnull(exhaust_fan_motor_temp_bpase) AND isnotnull(exhaust_fan_m
                    (exhaust fan motor temp c phase) AND ispotnull(motor side vibrations) AND ispotnull(exhaust motor side temp)
   | timechart span=120min avg(fan_side_vibrations) AS fan_side_vibrations_now , avg(chamber_b_temp_now, avg(chamber_b_temp_now, avg(chamber_a_temp) AS chamber_a_temp) AS chamber_a_temp  AS chamber_a_temp  As chamber_a_temp.
                    exhaust_fan_motor_temp_a_phase_now, avg(exhaust_fan_motor_temp_b_phase) AS exhaust_fan_motor_temp_b_phase_now, avg(exhaust_fan_motor_temp_c_phase) AS exhaust_fan_motor_temp_b_phase_now, avg(exhaust_fan_motor_temp_b_phase_now, avg(exhaust_fan_moto
                    avg(exhaust fan speed) AS exhaust fan speed now, avg(exhaust fan side temp) AS exhaust fan side temp now, avg(exhaust motor side temp) AS exhaust motor side temp) AS exhaust motor side temp) AS exhaust fan speed now, avg(exhaust fan speed) AS exhaust fan speed now, avg(exhaust fan 
                    | where NOT (day_of_week="Saturday" OR day_of_week="Sunday")
                         fields + motor side vibrations fan side vibrations chamber b temp chamber a temp exhaust fan motor temp a phase exhaust fan motor temp b phase exhaust fan motor temp c phase
                          where isnotnull(exhaust_fan_motor_temp_b_phase) AND isnotnull(chamber_b_temp) AND isnotnull(chamber_b_temp) AND isnotnull(exhaust_fan_motor_temp_a_phase) AND
                                             snotnull(exhaust_fan_motor_temp_c_phase) AND isnotnull(motor_side_vibrations) AND isnotnull(exhaust_motor_side_temp)
                    | timechart span=120min avg(chamber_a_temp) AS chamber_a_temp_later24
       timechart span=2h avg(*) AS *_now avg(*_later24) AS *_later24
```



#### Use Case #2

#### Try to find anomalies in motor features

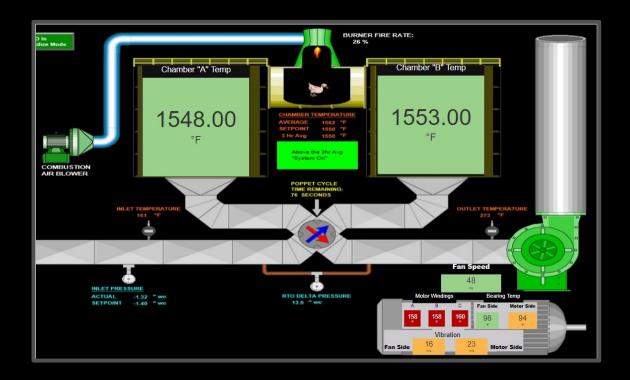
- Using Density we chose features that should be tied to motor health. This does not necessarily predict temperature but it might predict a motor failure.
- Each feature does have anomalies everyday, at what point do the number of anomalies predict an issue?



#### Use Case #3

How to display the information for the engineers

The equipment engineers have a system to view the status of the RTO and its features. But their system does not trend, display detailed information, or alert via email or text. But we took the visual and created a glass table to allow this information to be displayed in a useful way. Each colored field can be clicked which takes you to a trend report for that data.



#### Results

#### Has this resulted in any positive impact?

- ▶ The RTO has not failed since we began. But the engineers have been alerted by looking at the Glass Table that the motor vibrations were abnormal. It turned out that preventive maintenance had been preformed and an adjustment needed to be made. One of the phases also had a higher than normal temperature and that prompted an investigation. The ambient temperature was higher than normal in the area and caused the temperature to be higher.
- ▶ Also, between shifts one of the burners went out and the temperature dropped below the alert temperature. Since this was between shifts there was no danger of pollutants escaping. But, we have alerts setup to monitor the current temperature of the chambers. This emailed and texted the engineers. Their system sounds an alarm on the PLC, but since it was between shifts no one was around to hear. This saved an hour of downtime for the plant. Because the engineers were able to respond to the alert and replace the burner before 2<sup>nd</sup> shift started.

#### What have I learned?

#1

Machine learning is everywhere: Netflix recommendations, Twitter Sentiment Analysis for businesses and political figures, Autonomous Cars, and Video Games. #2

ML is easier and harder than I thought. The math behind machine learning is extremely complex and best left to PhDs. But using the services and platforms that leverage this math is way easier than the math.

#3

Gathering and cleaning data for ML is just as much or more work than actually building ML models.

#4

C-Level bosses like to hear Machine Learning but that is all they want to hear. No more details. #5

99.9% of people will be intimidated by the words machine learning. It is up to you as the practitioner to use visualizations and examples to help your customers understand the benefits it could provide.



#### **Bonus Tip!**

#### **Convince Your Boss**

- What to do when your boss expresses doubt about machine learning:
  - Have Brian Gilmore come give a demo of the MLTK!
  - If his schedule is full, then you walk through the examples provided by the MLTK.
  - Reference other Splunk .conf sessions, I attended a more than a few during the previous .conf's. And they are all available online at Splunk's .conf website.
  - Have an a couple of ideas that can provide benefit (ROI) to your company. Reduce power usage, employee attendance, equipment breakdown, cybersecurity (UBA- user behavior analysis). If you can show you have real use cases and ROI it becomes much easier to get your boss to approve a ML project.

#### Where are we going now?

- After the help provided by the MLAC we are applying the knowledge gained to other areas:
  - We have recently brought in all the lift motors in our Frame Assembly plant. We plan on applying the same techniques to help avoid downtime.
  - We have been collecting power, chill water, natural gas, and local weather data. I am working
    on how each of those interact and impact other factors. Is plant temperature tied to more or
    less defects per VIN? Etc.
  - Brake Fill and HAC Fill (air conditioning) data, and correlating how dew points impact the measuring of Torr (amount of vacuum) in each.
  - Our team and the Weld team will begin evaluating their lift motors to determine which motors to replace first in a department refresh. Again using the techniques we learned to determine the least healthy motors.

## "Success represents the 1% of your work which results from the 99% that is called failure"

Soichiro Honda

#### Q&A

Charlie Studdard | Machine Data Analytics





## conf19 splunk>

## Thank You!

Go to the .conf19 mobile app to

RATE THIS SESSION