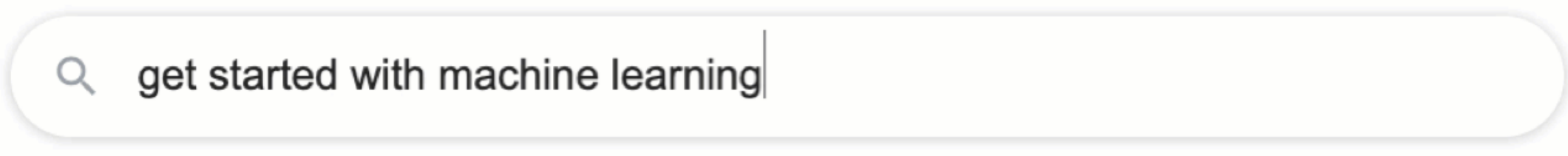
Machine Learning

Proven Applications and New Features

Seth DeLand

MathWorks

How to Get Started with Machine Learning?

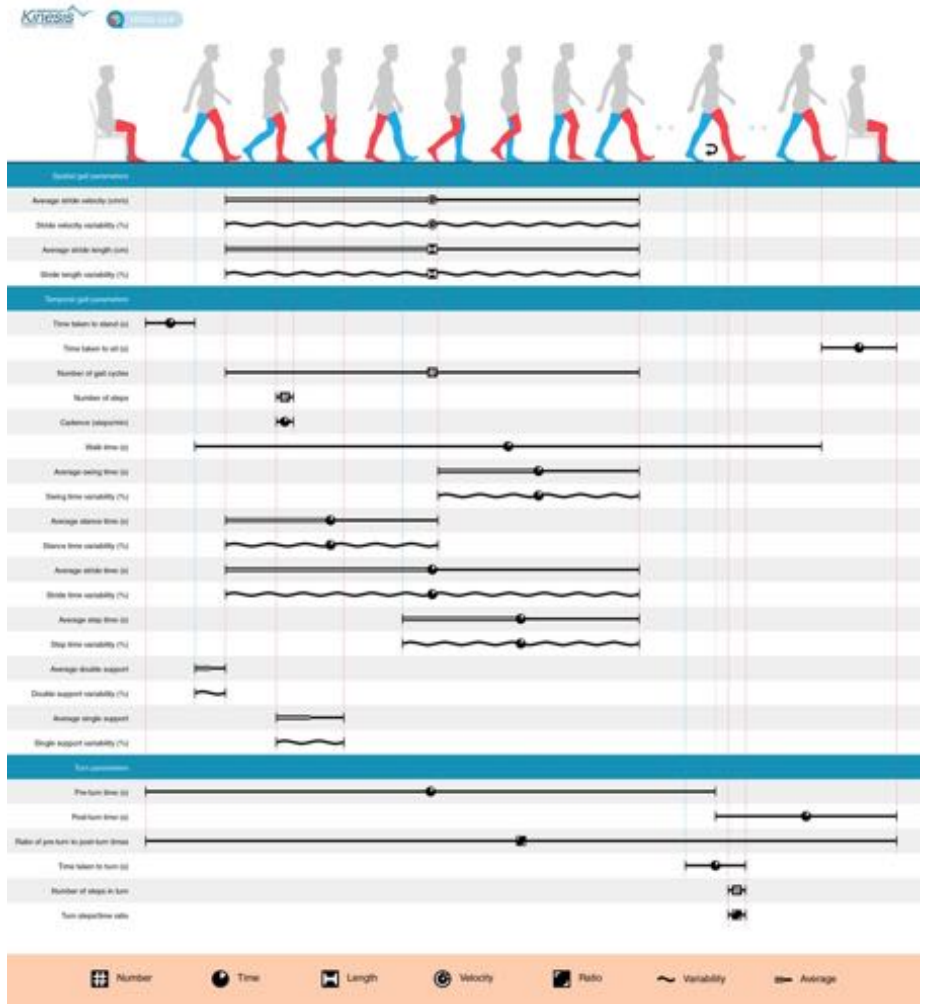


About 611,000,000 results (0.63 seconds)

Machine Learning Success Stories

Kinesis Health Technologies

Predicting a patient's fall risk with machine learning.



Machine Learning

+

X

Machine Learning

+

Industry Knowledge

Application Knowledge

Your Own Expertise

Examples of Successful Machine Learning Applications

Fleet Data Analytics

Energy Forecasting

Manufacturing Analytics

New Capabilities

- MATLAB apps
- AutoML
- Signal Processing with Machine Learning
- C/C++ Code Generation

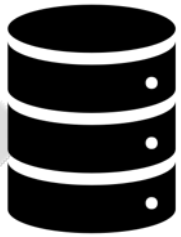
Examples of Successful Machine Learning Applications

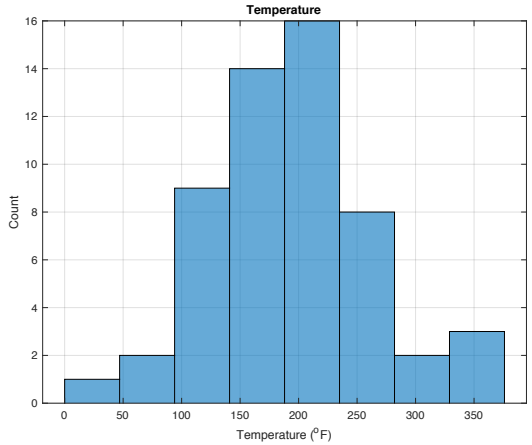
Fleet Data Analytics

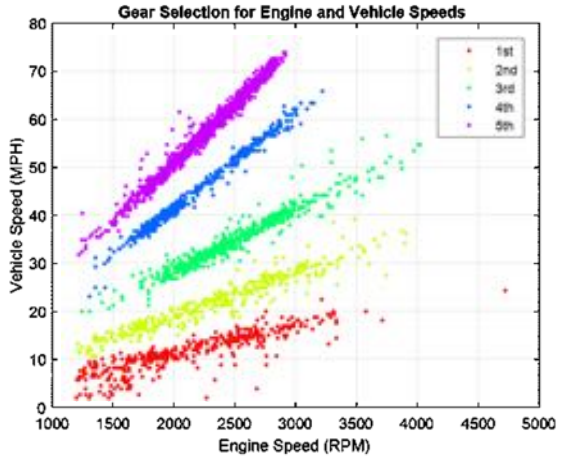
Energy Forecasting

Manufacturing Analytics

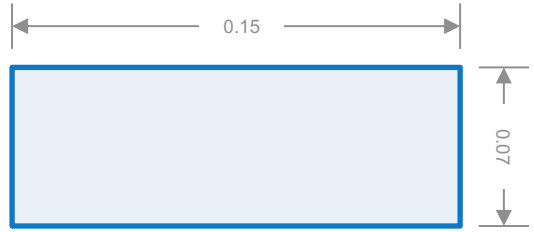
Fleet Data Analytics



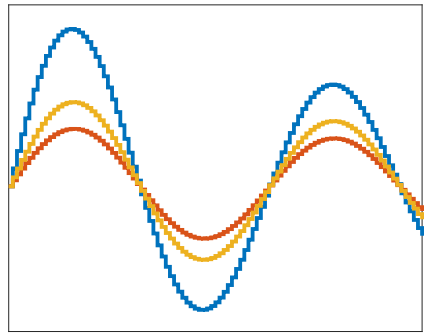




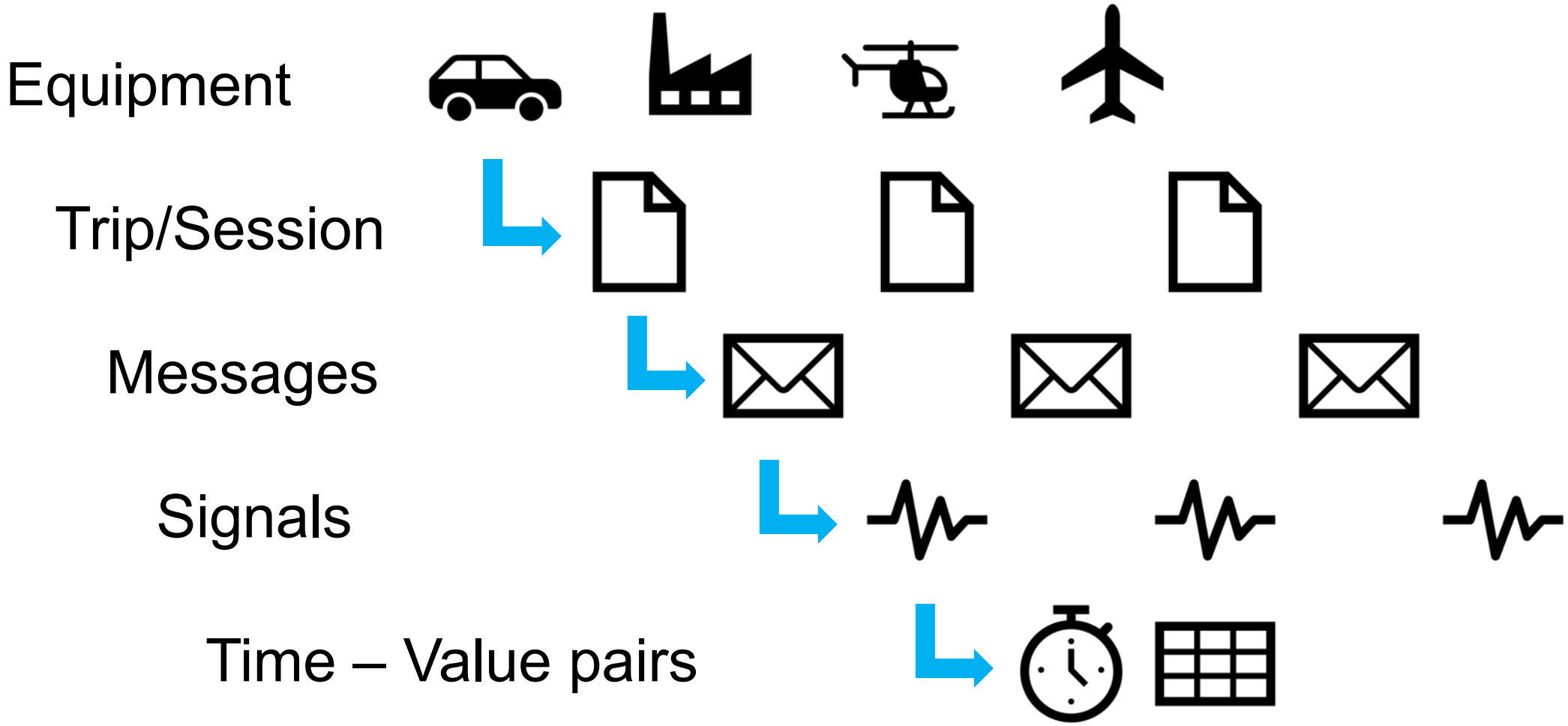
Design Decisions



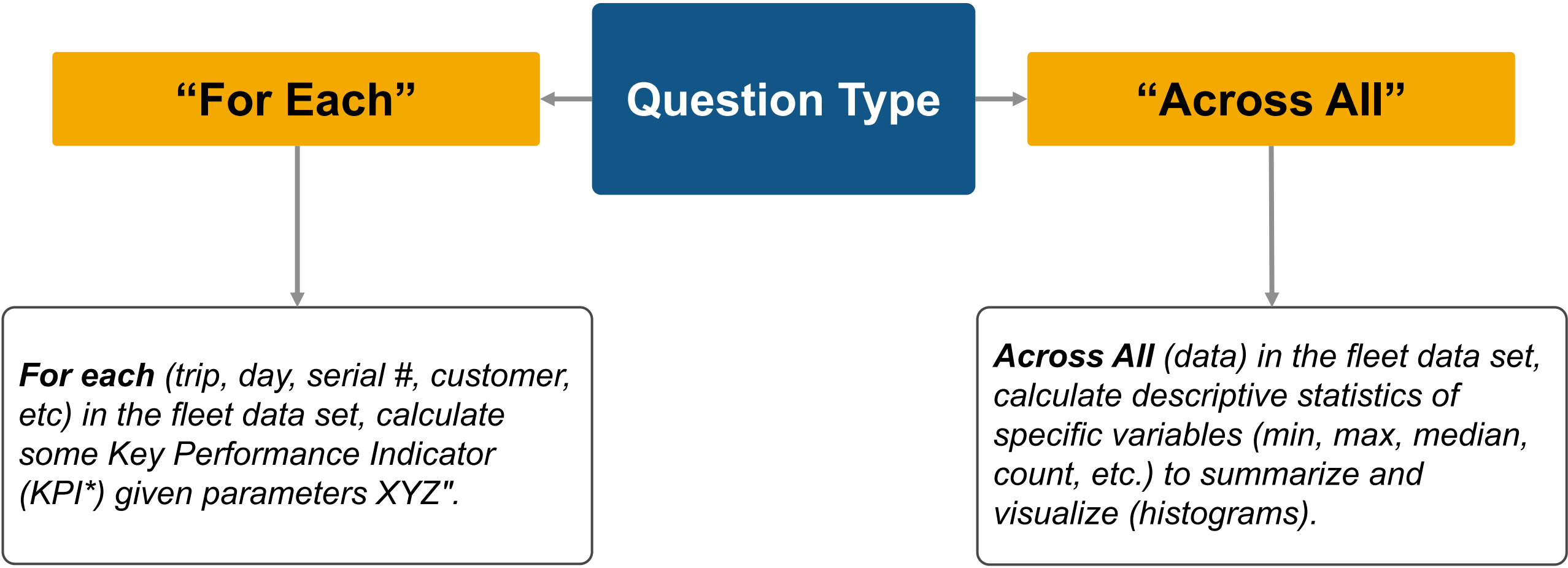
Test Plans



What Level of Data?



What Type of Question?



Scale to Large Collections of Data with Datastore

Create a datastore from all CSV files

```
ds = datastore('*.csv')
```

Read a single file of data

```
data = read(ds);
```

Reset the datastore back to the first file

```
reset(ds);
```

Find the maximum value of "Y" in each file

```
X = [];
while hasdata(ds)
        data = read(ds);
        X(end+1) = max(data.Y);
end
```

Available Datastores	
General	`datastore`
	`spreadsheetDatastore`
	`tabularTextDatastore`
	`fileDatastore`
Database	`databaseDatastore`
Image	`imageDatastore`
	`denoisingImageDatastore`
	`randomPatchExtractionDatastore`
	`pixelLabelDatastore`
	`augmentedImageDatastore`
Audio	`audioDatastore`
Predictive Maintenance	`fileEnsembleDatastore`
	`simulationEnsembleDatastore`
Simulink	`SimulationDatastore`
Automotive	`mdfDatastore`
Custom	subclass `matlab.io.Datastore`
Transformed	`transform` an existing datastore

Performing “Across All” Calculations with Tall

Create a datastore from a collection of CSV files, and select the "Time" and "EngineSpeedRPM" variables.

```
ds = datastore('EngineData*.csv',...
    "SelectedVariableNames",["Time","EngineSpeedRPM"]);
```

Create tall table:

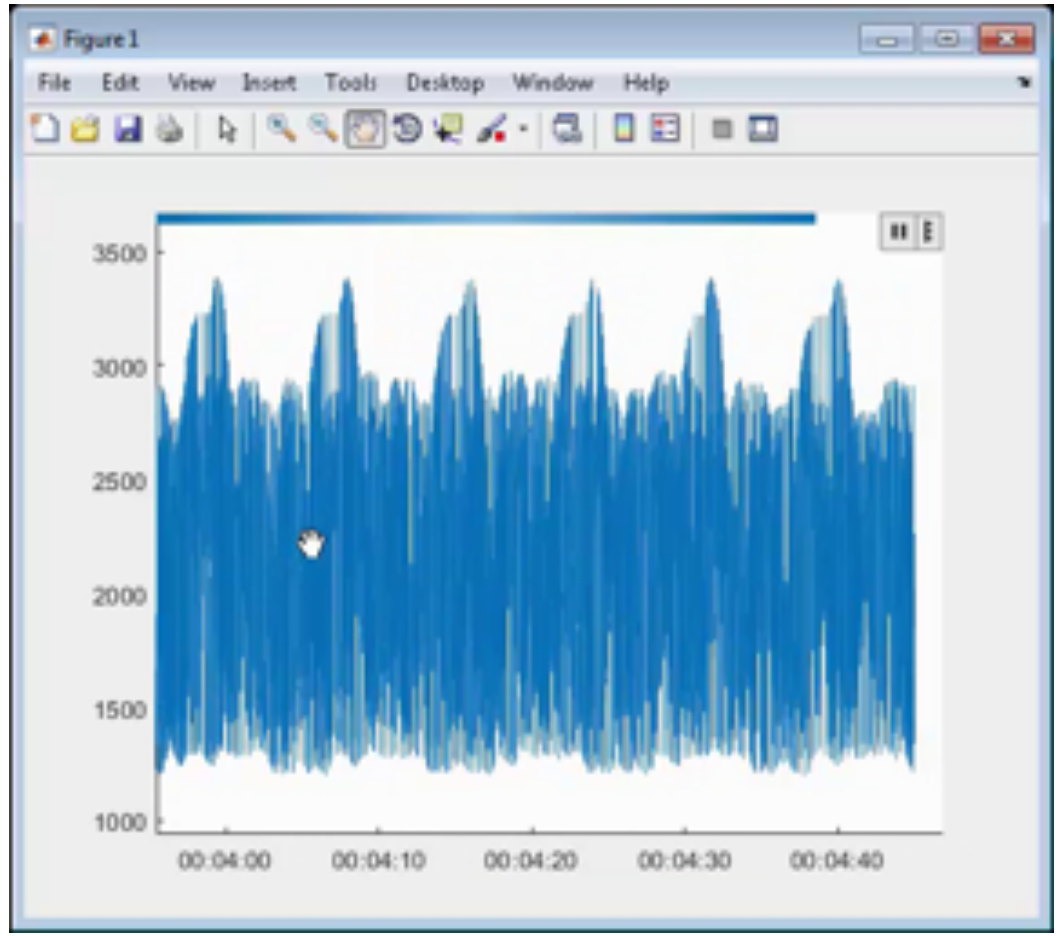
```
t = tall(ds);
```

Convert to tall timetable:

```
tt = table2timetable(t);
```

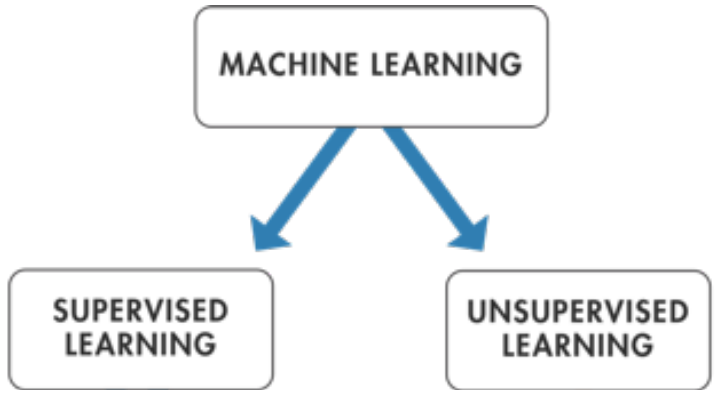
Plot EngineSpeedRPM vs. Time:

```
plot(tt.Time,tt.EngineSpeedRPM)
```



- Visualizations
- Data preprocessing
- Machine Learning

Exploring Fleet Data with Unsupervised Learning



Unsupervised Learning for Operational Mode Clustering

Plot the raw data:

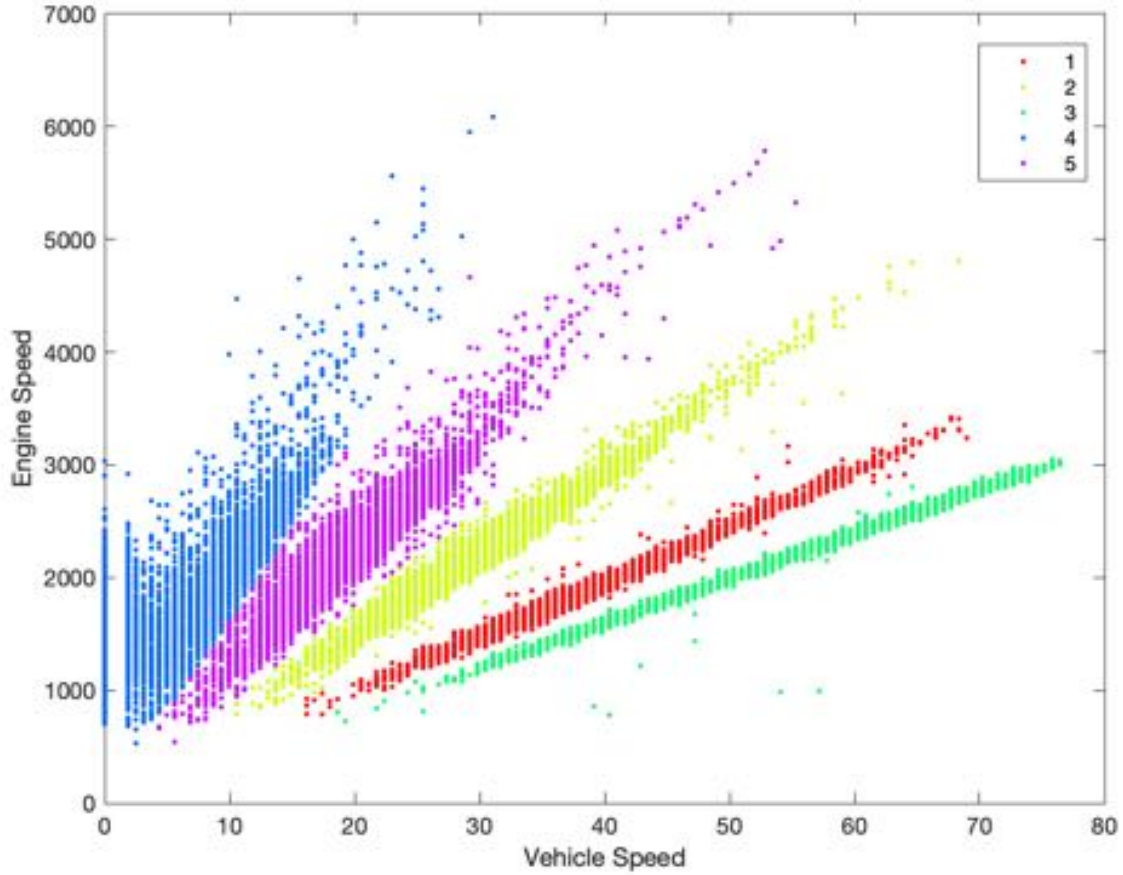
```
figure;
plot(t.Speed_OBD_,t.EngineRPM,'.k')
xlabel('Vehicle Speed');
ylabel('Engine Speed');
```

Cluster the data with the K-Means algorithm:

```
X = [t.Speed_OBD_,t.EngineRPM];
IDX = kmeans(X,5,"Distance","cosine");
```

Plot results of the clustering:

```
gscatter(t.Speed_OBD_,t.EngineRPM,IDX);
xlabel('Vehicle Speed');
ylabel('Engine Speed');
```



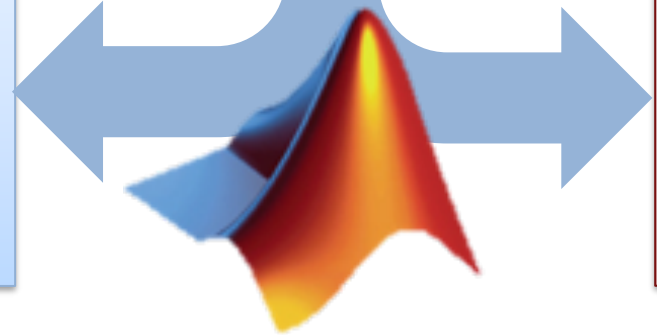
Deploying Fleet Analytics

Vehicle data, driver profiles

“Cold Storage”

Historic data:

- **Batch processing**
- Large data on cluster
- Explore long term trends
- Build models



“Hot Storage”

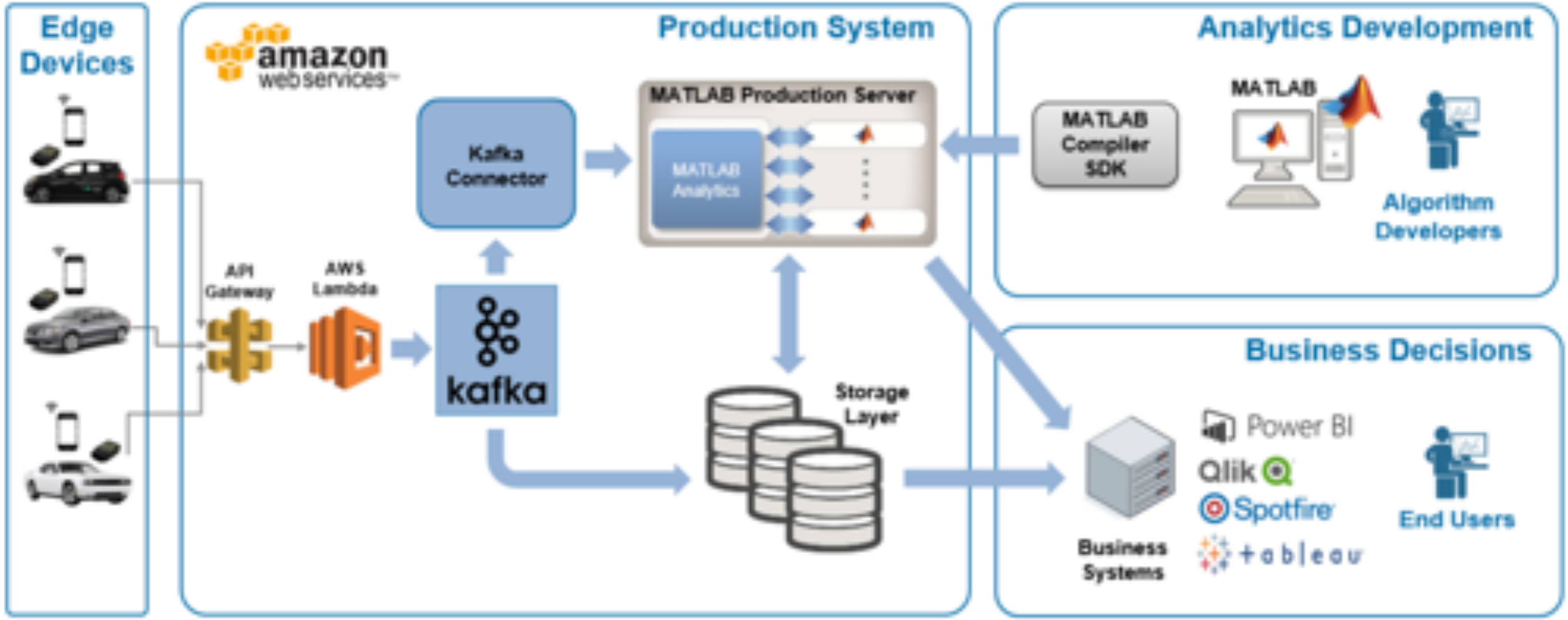
Streaming data:

- **Near real-time**
- Test and implement model for new data
- Stream processing

Spark

kafka

Fleet Analytics Streaming Architecture



Fleet Analytics in Practice: Volkswagen Data Lab

Develop technology building block for tailoring car features and services to individual

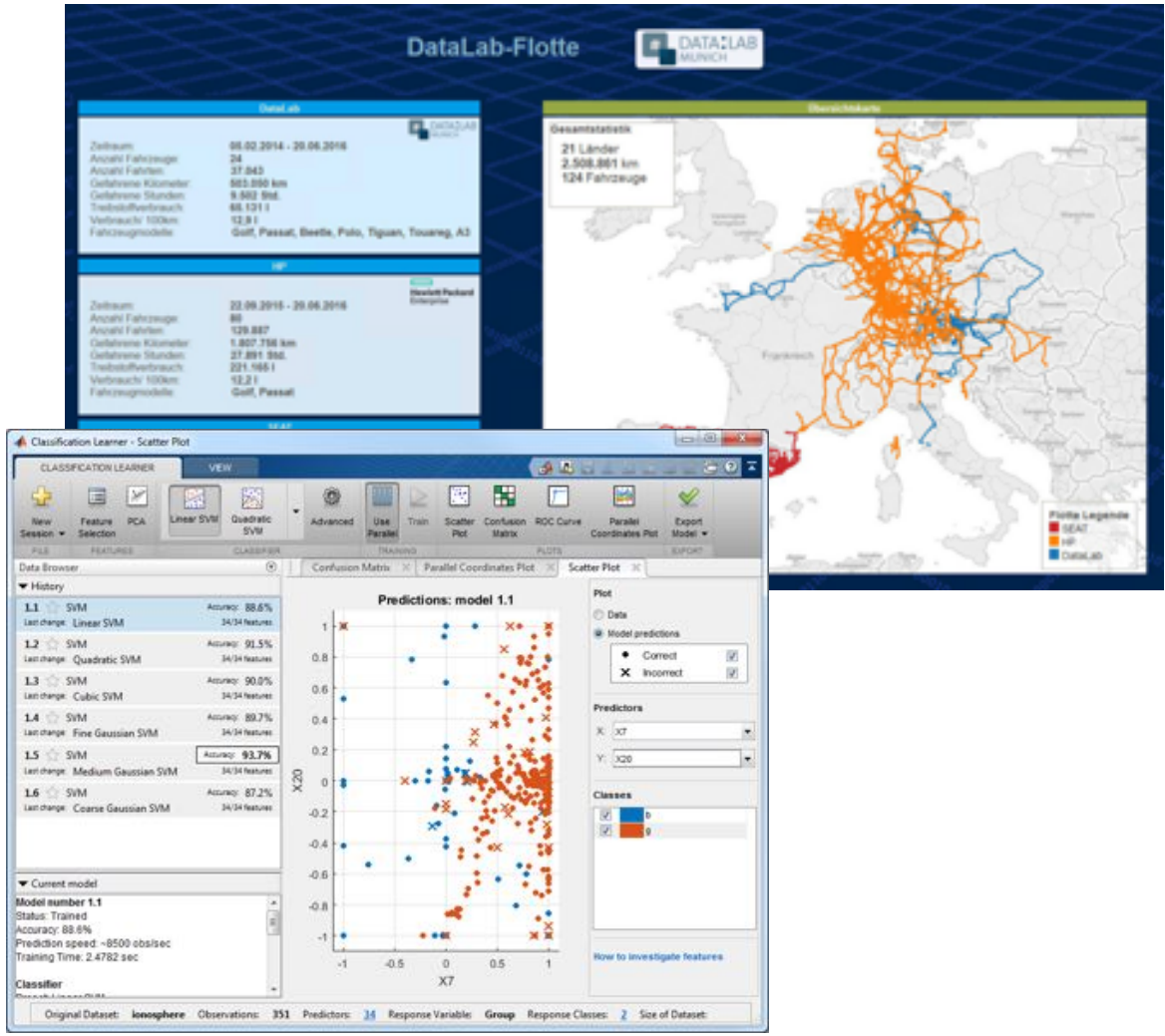
- Driver and Fleet Safety
- Driver Coaching
- Driver-Specific Insurance

Data sources

- Logged CAN bus data and travel record

Results

- Proof-of-concept model for “telematic fingerprint”
- Basis for the “pay-as-you-drive” concept



Source: “Connected Car – Fahrererkennung mit MATLAB“
Julia Fumbarev, Volkswagen Data Lab
MATLAB EXPO Germany, June 27, 2017, Munich Germany

Machine Learning + X

Fleet Analytics

Equipment Expertise

Design Specs

Operating Modes

Operating Conditions

Machine Learning

Statistical Analysis

Unsupervised Learning

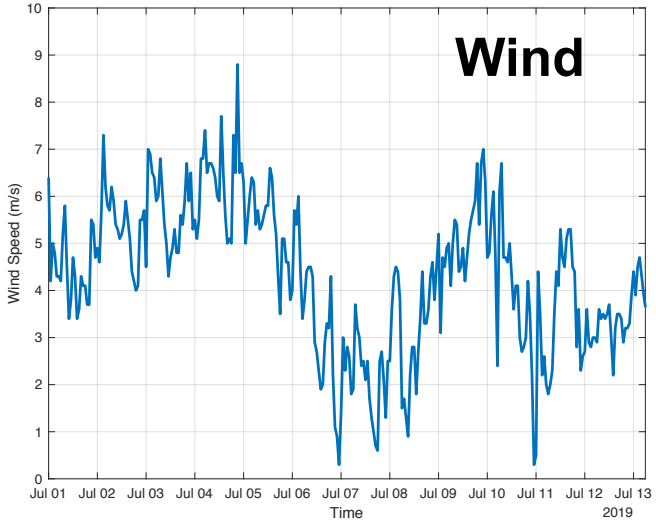
Examples of Successful Machine Learning Applications

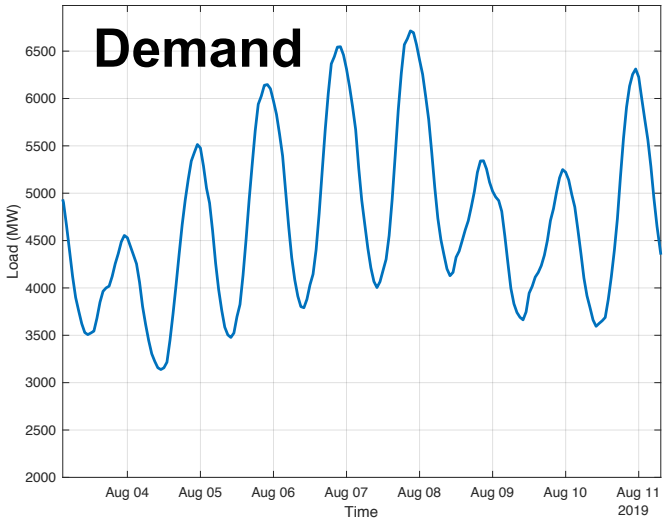
Fleet Data Analytics

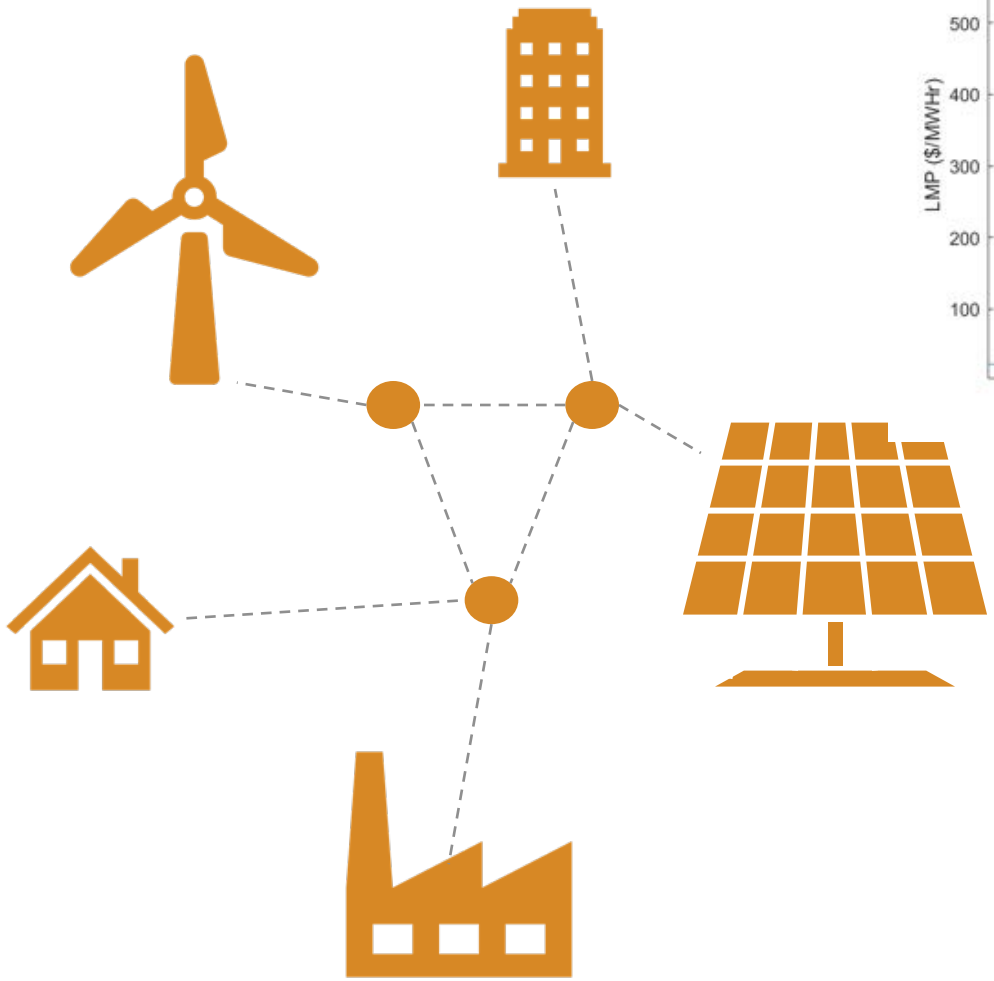
Energy Forecasting

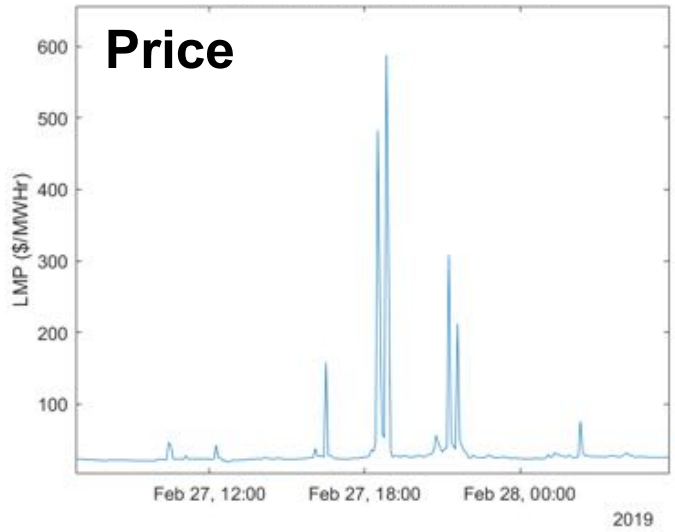
Manufacturing Analytics

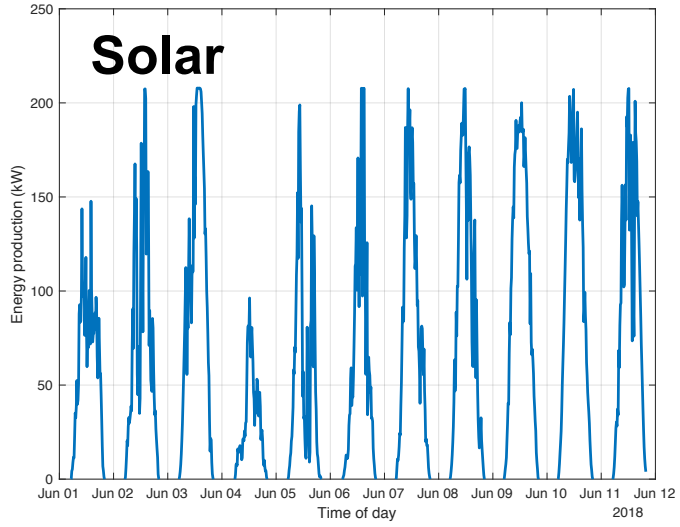
The Need for Energy Forecasts







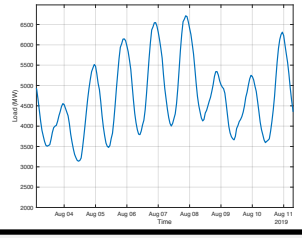




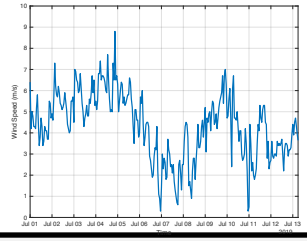
How Energy Forecasting Works

Historical Data

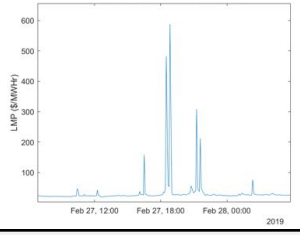
Electricity Demand



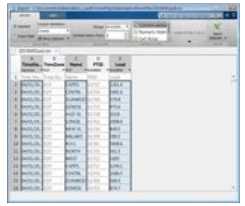
Weather



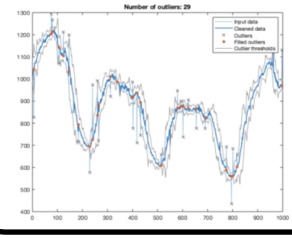
Electricity Prices



Combine



Preprocess



Features

load *wind*

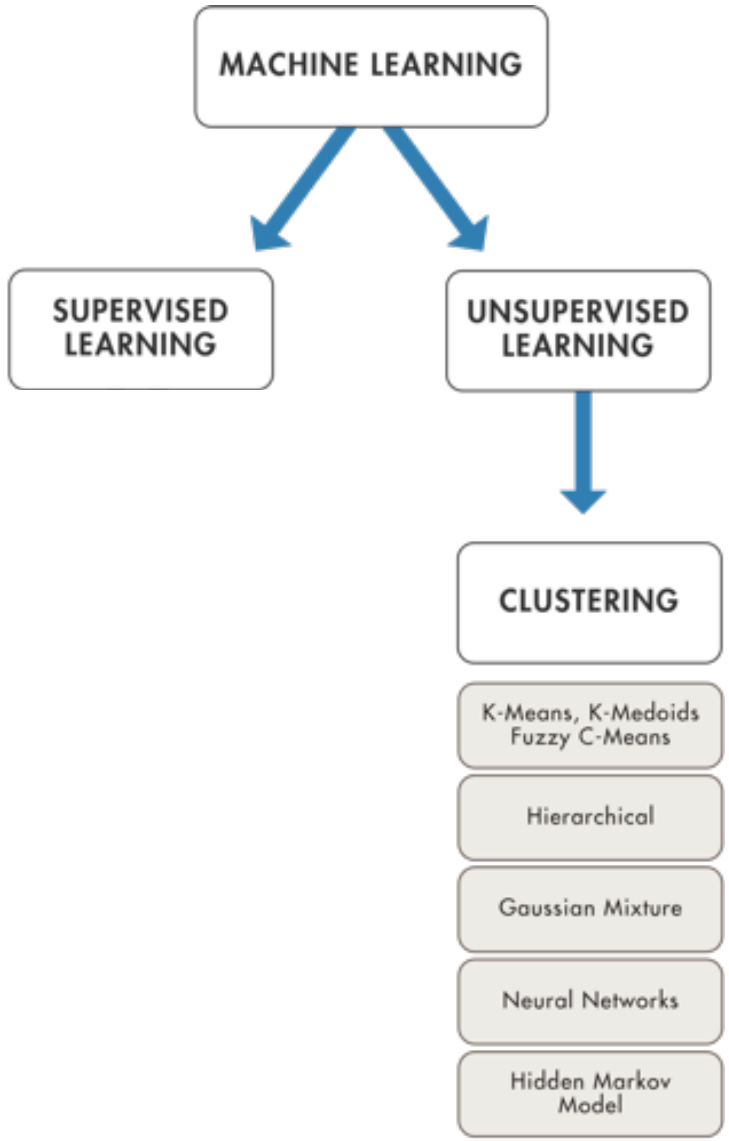
temp *24hr*

day *1week*

month

Machine Learning

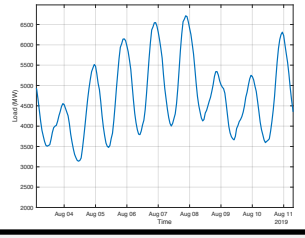
Building Forecast Models with Regression Techniques



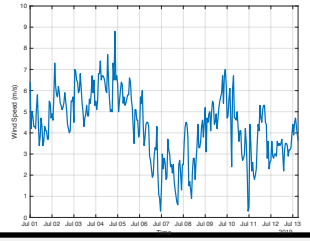
Using Energy Forecasting Models

New Data

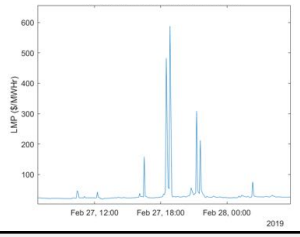
Electricity Demand



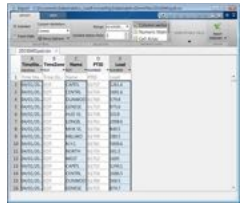
Weather



Electricity Prices



Combine



Features

load *wind*

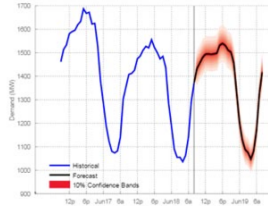
temp *24hr*

day *1week*

month

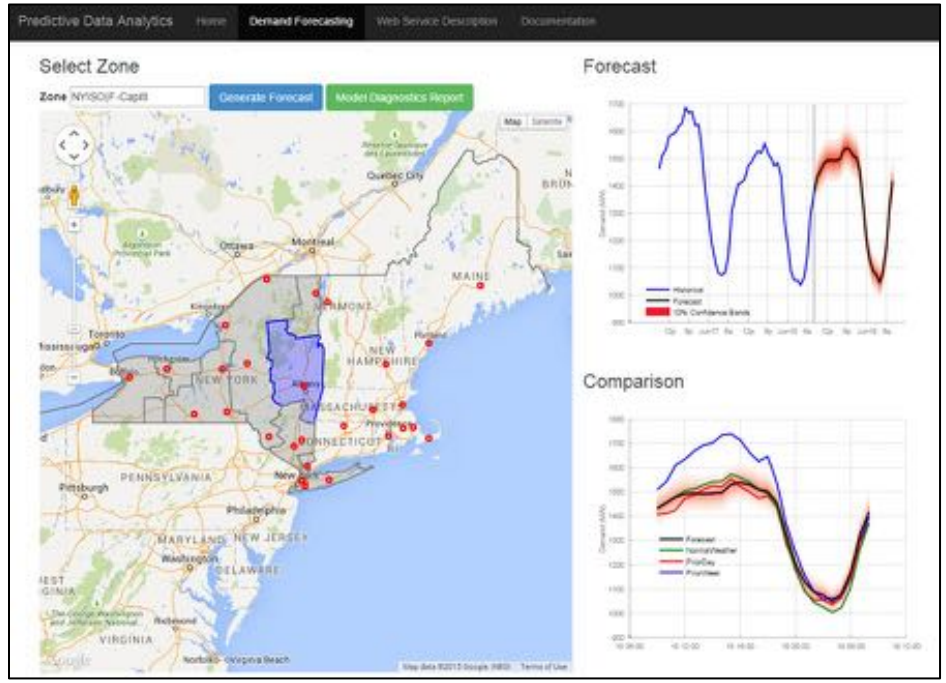
Trained Machine Learning Model

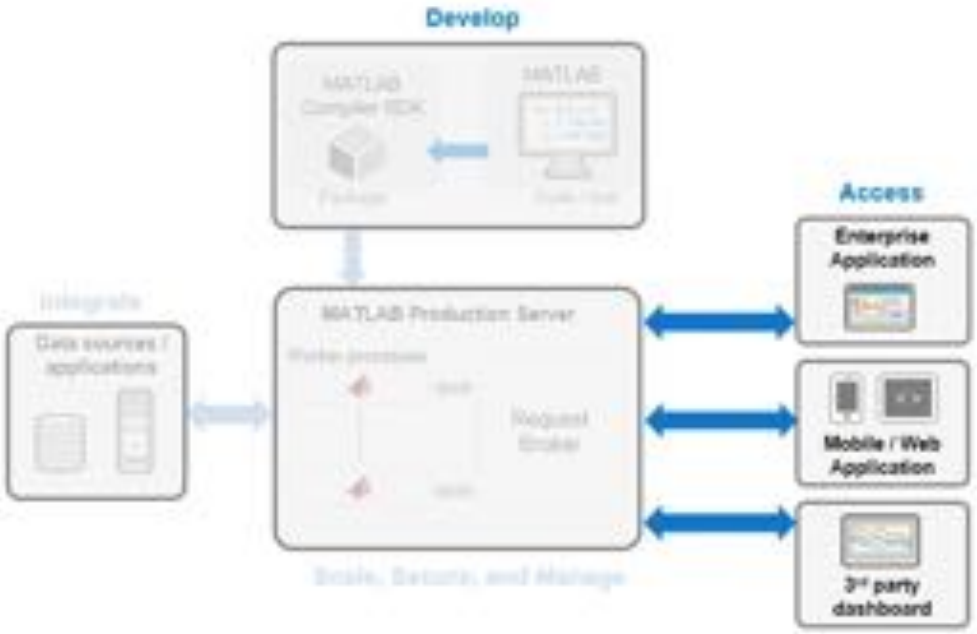
Forecast



Deploying Energy Forecasts

Dashboards for operators and traders





API for App Developers

Combining Forecasting with Optimization

"When should I operate my generators to maximize the return on my investment?"

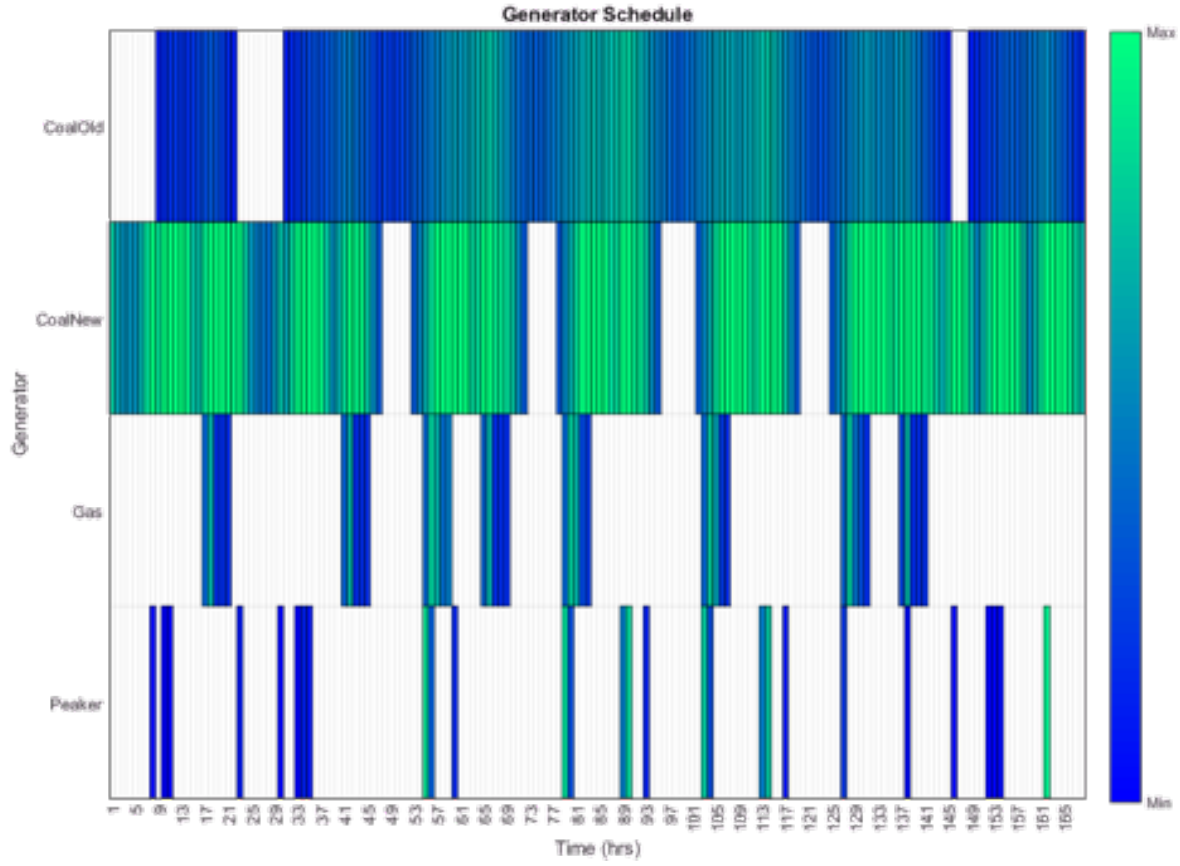
Optimization Problem:

Minimize:

Cost of generating electricity

Constraints:

1) *Meet forecasted demand*
2) *Operational constraints*
3) *Etc.*



Energy Forecasting in Practice: Naturgy Energy Group S.A.

Challenge

Maximize margins in energy trading by predicting available supply and peak demand

Solution

Use MATLAB to build and optimize models that incorporate historical data, weather forecasts, and regulatory rules

Results

- Response time reduced by months
- Productivity doubled
- Program maintenance simplified

Portomouros hydroelectric dam.

"Because we need to rapidly respond to shifting production constraints and changing demands, we cannot depend on closed or proprietary solutions. With MathWorks tools we get more accurate results — and we have the flexibility to develop, update, and optimize our models in response to changing needs."

- Angel Caballero, Gas Natural Fenosa

Link to user story

Machine Learning + X

Fleet Analytics

Equipment Expertise

Design Specs

Operating Modes

Operating Conditions

Machine Learning

Statistical Analysis

Unsupervised Learning

Energy Forecasting

Electrical Grid Expertise

Seasonality

Weather Effects

Generator Characteristics

Machine Learning

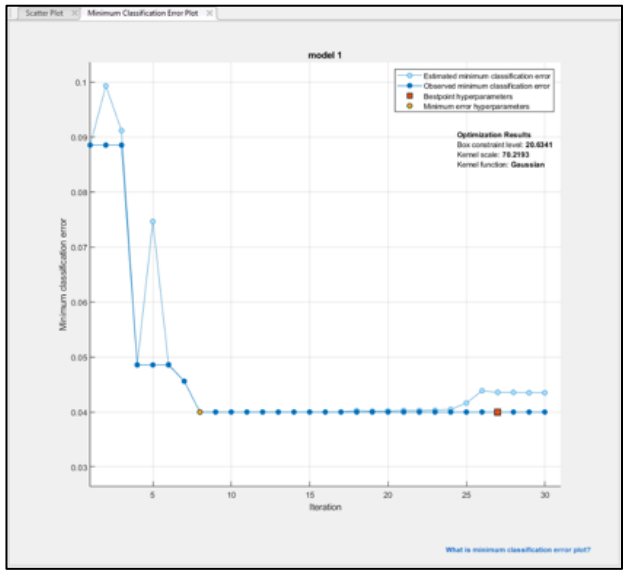
Time Series Modeling

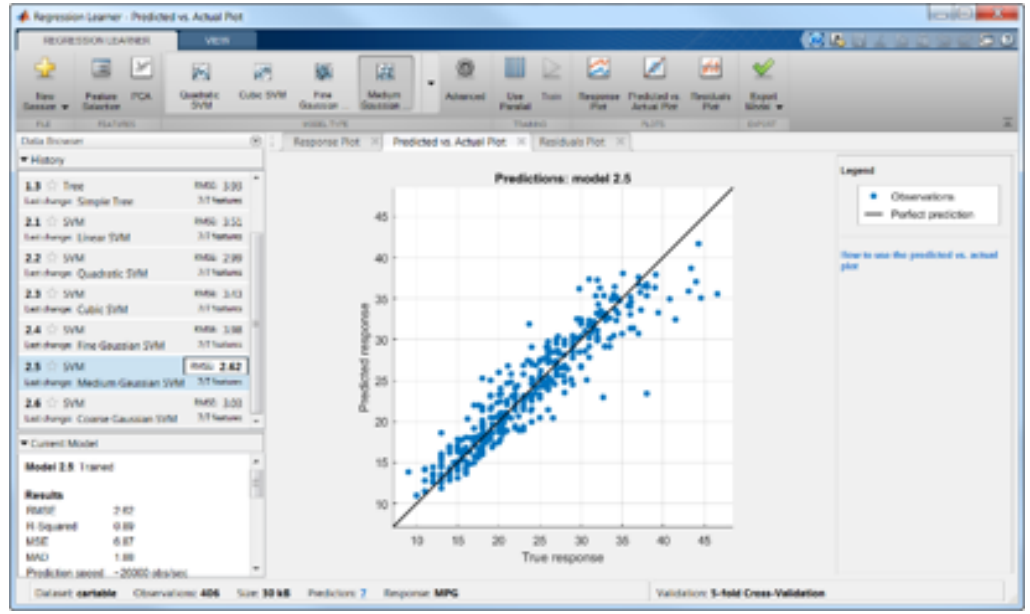
Regression

Machine Learning apps

- Try out many models
- Compare Results
- Get to a reasonable model without worrying about the details

Perform Hyperparameter Optimization in apps





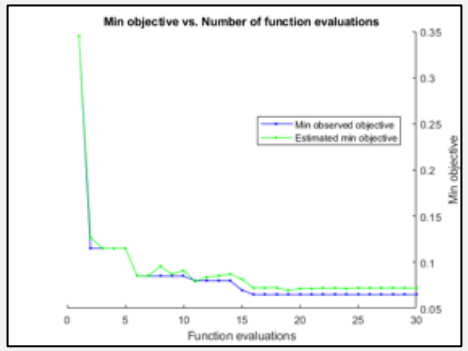
AutoML

- Build many machine learning models
- Find a good model without becoming an expert

Model Selection

`fitcauto`

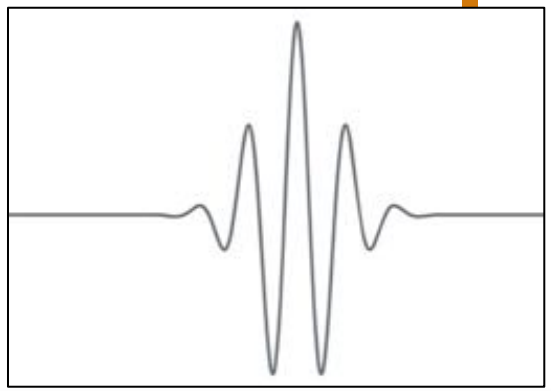
Hyper-parameter Optimization

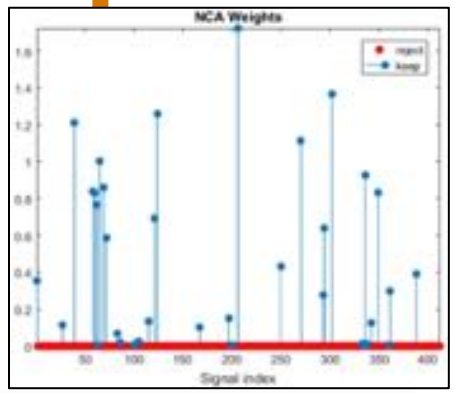


Decision Tree?
SVM?
KNN?
Ensemble?
…?

Import Data → **Preprocess Data** → **Extract Features** → **Train Model** → **Deploy & Integrate**

Wavelet Scattering





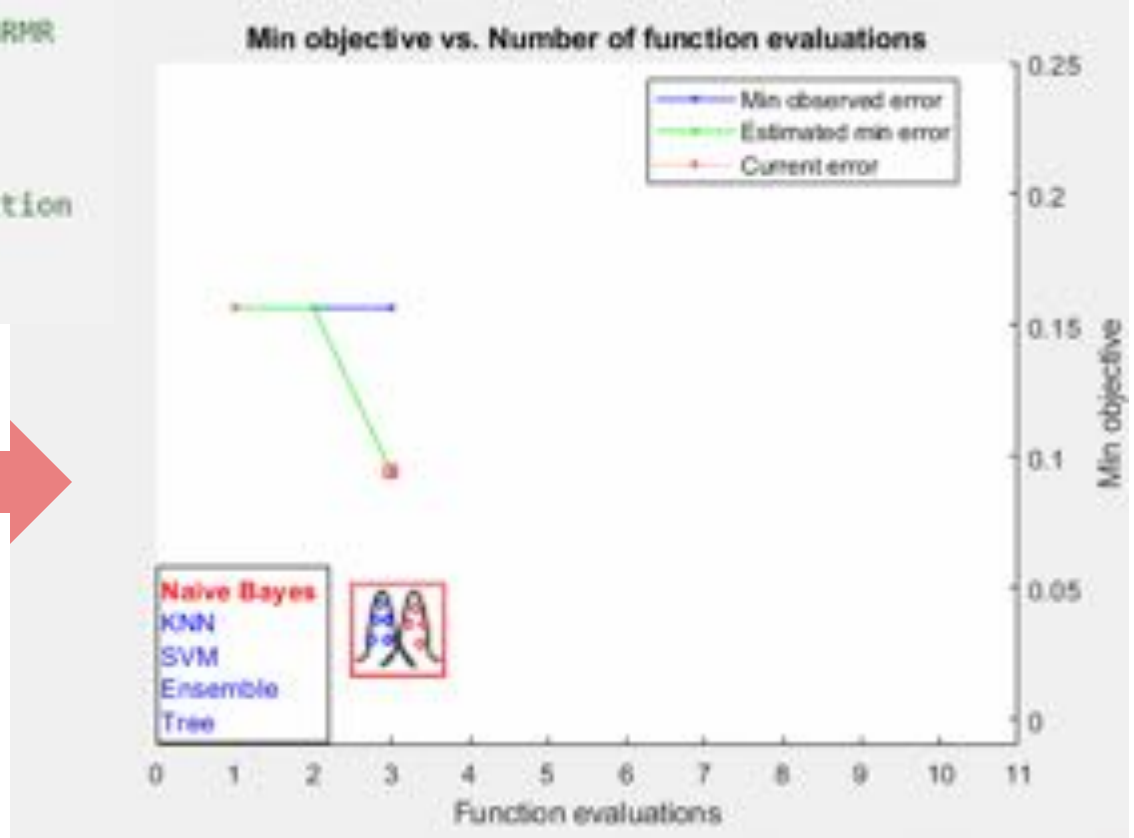
Feature Selection

AutoML “in action”

```
% Step 1: apply Wavelet scattering to extract features
sf = waveletScattering('SignalLength',N, 'SamplingFrequency',50);
Wfeatures = featureMatrix(sf,thisSignal(1:N),'Transform','Log');
% do this across signals <thisSignal> and accumulate <allFeatures> with labels

% Step 2: select top <featN> features according to feature ranking, e.g. MRMR
[mrmrFeatures , scores] = fscmrmr(allFeatures, 'class');
trainFeatures = allFeatures(:, [mrmrFeatures(1:numPredictorsToUse);true]);

% Step 3: Select optimized model from 100 iterations of 1-step model selection
modelAuto = fitcauto(trainFeatures,'class', 'Learners','all',
'MaxObjectiveEvaluations',100);
```



Examples of Successful Machine Learning Applications

Fleet Data Analytics

Energy Forecasting

Manufacturing Analytics

What is Manufacturing Analytics?

Definition: Apply modeling (**AI**) to **process** and **sensor data** to maximize operational performance

Key Use Cases:

1. **Automate** the **monitoring** of manufacturing process
2. Ensure **product quality**
3. **Optimize yield** of complex production processes

Challenges in Applying AI to Manufacturing

Lots of Data – much in “Data Historians” (SCADA, LIMS, OSISoft PI)

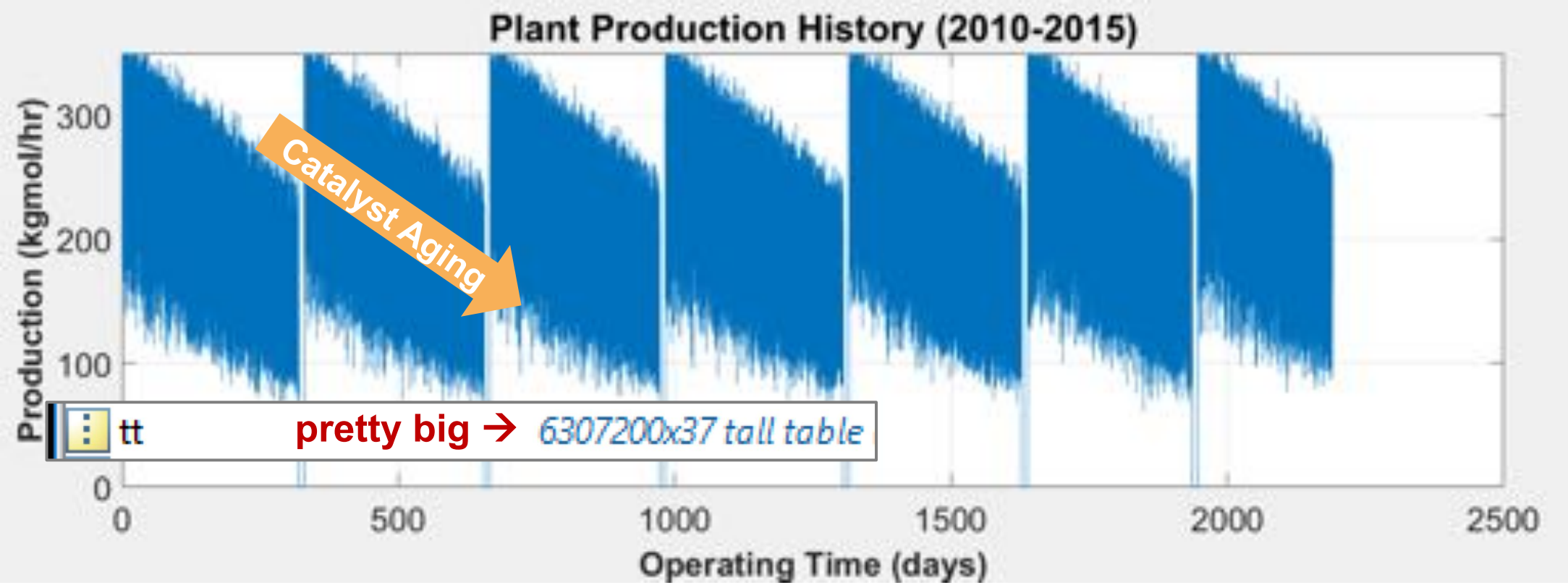
Reliable measurements or modeling

- Sensor failures
- Hidden variables

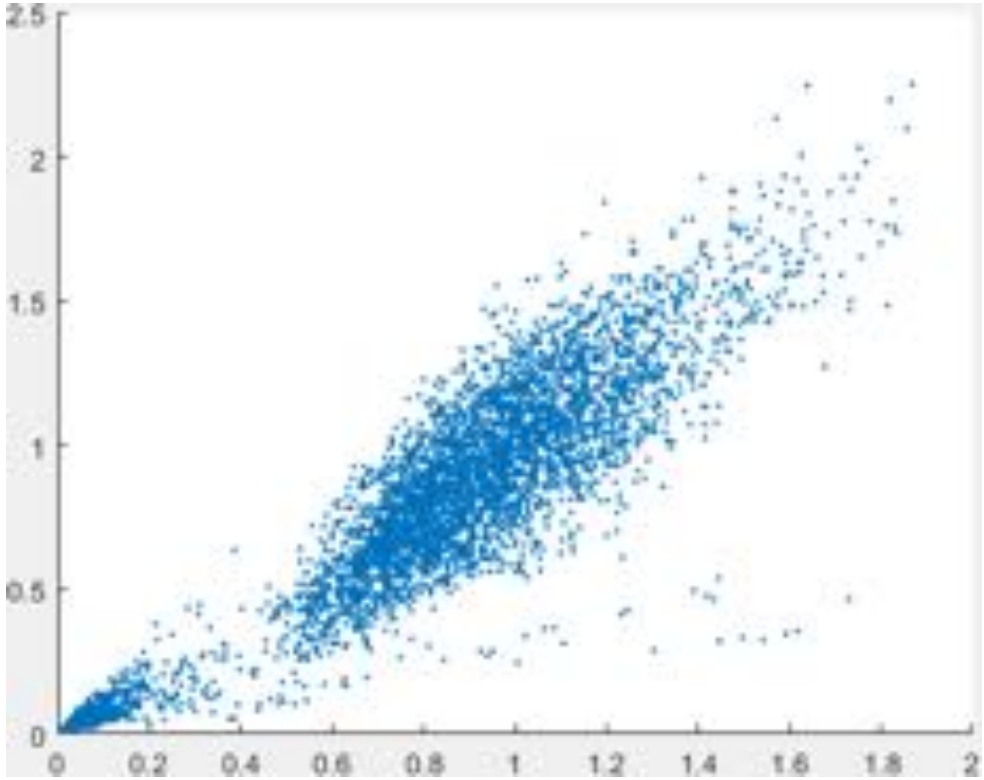
Use of many different tools

- Limited Predictive modeling
- Handle streaming data
- Customization

Uncover Hidden Variables with Process Modeling

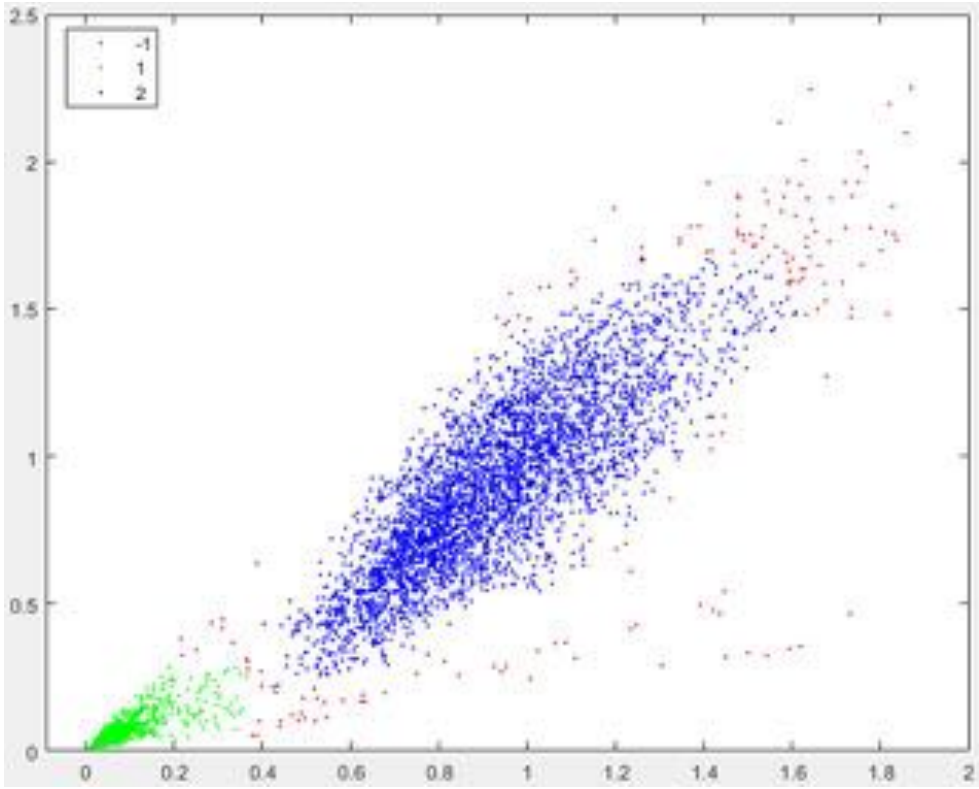


Case Study: Anomaly Detection

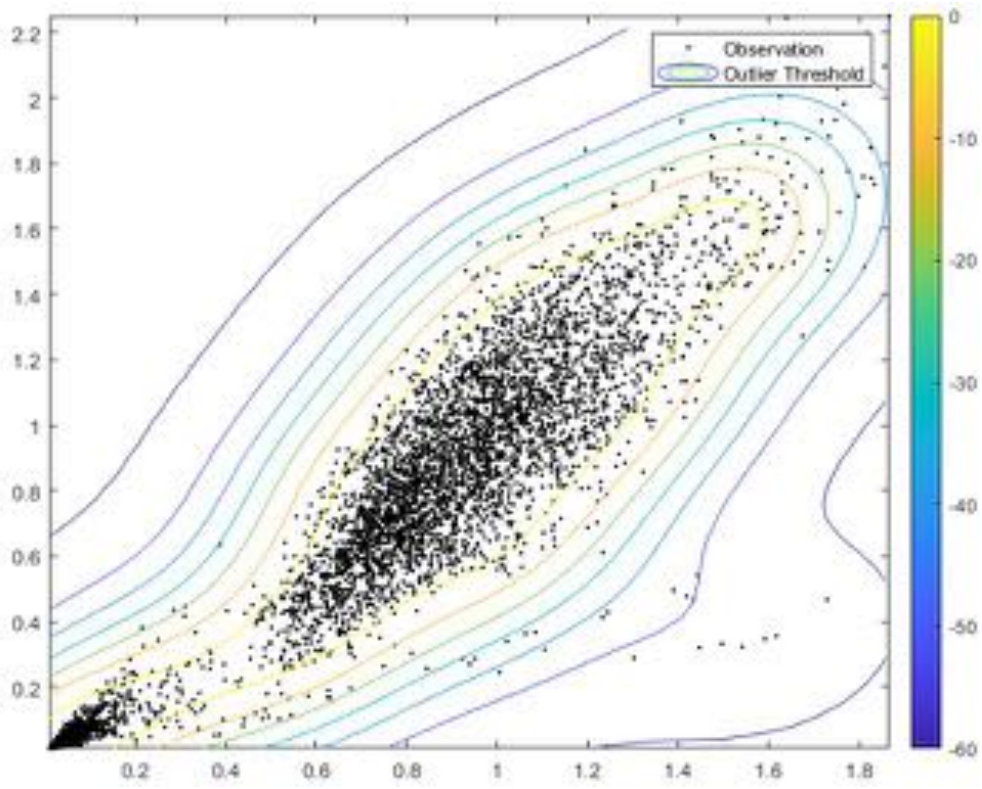


Case Study: Anomaly Detection

1. Cluster with **DBSCAN**



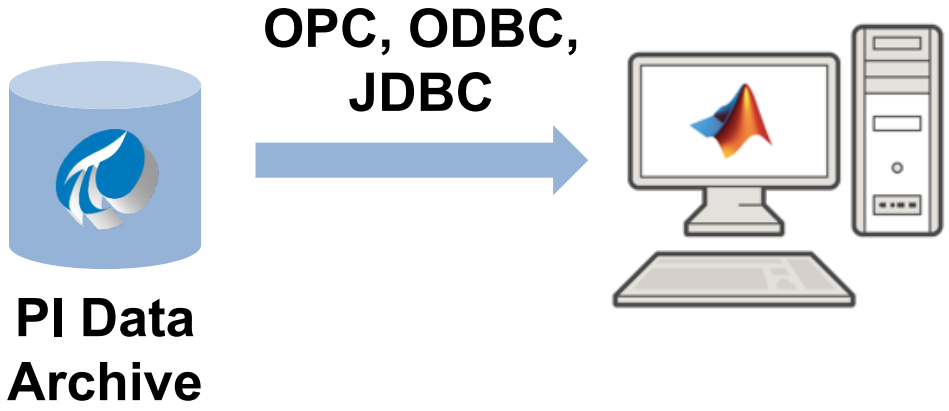
2. One-class SVM

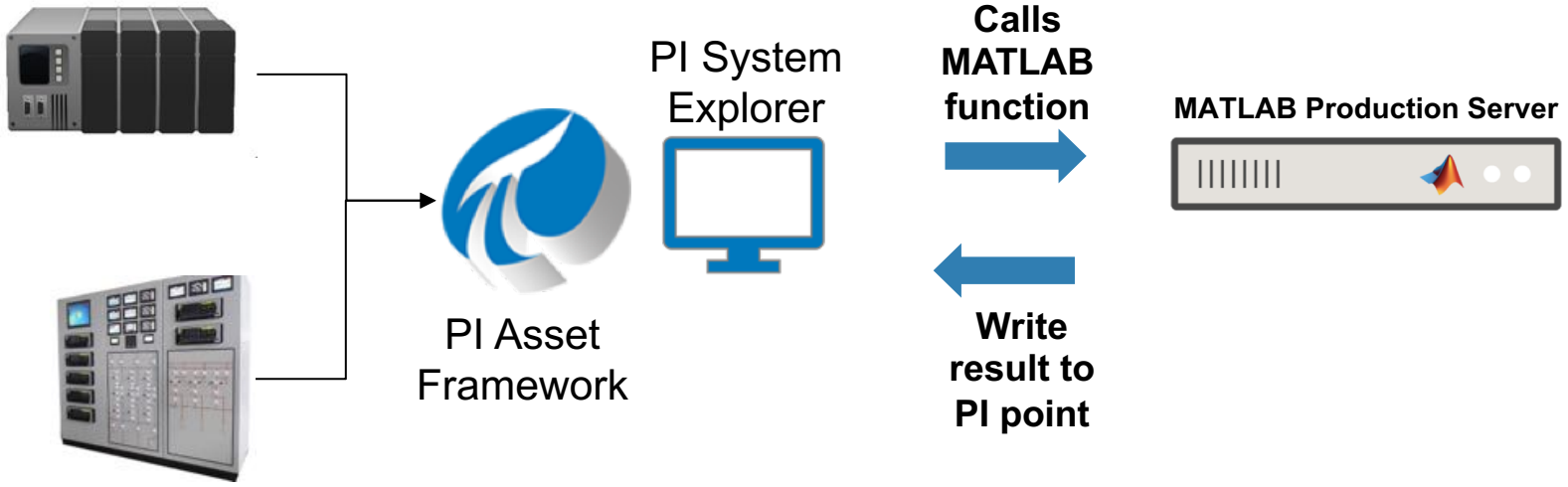


Deployment

Integration with Data Historians

- OPC Toolbox (Database tbx via ODBC or JDBC) connects with PI Server





Customize Analytics Delivery

- Accessing insights via GUI critical for plant staff and process engineers
- Build a custom dashboard with App Designer

Machine Learning + X

Fleet Analytics

Equipment Expertise

Design Specs

Operating Modes

Operating Conditions

Machine Learning

Statistical Analysis

Unsupervised Learning

Energy Forecasting

Electrical Grid Expertise

Seasonality

Weather Effects

Generator Characteristics

Machine Learning

Time Series Modeling

Regression

Manufacturing Analytics

Manufacturing Expertise

Process Equipment

Variables & Set Points

Parameter Impact

Machine Learning

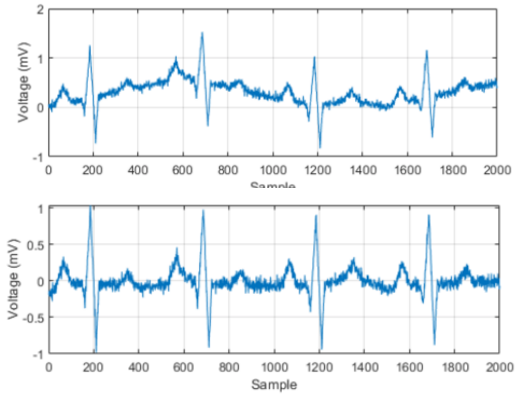
Anomaly Detection

Regression

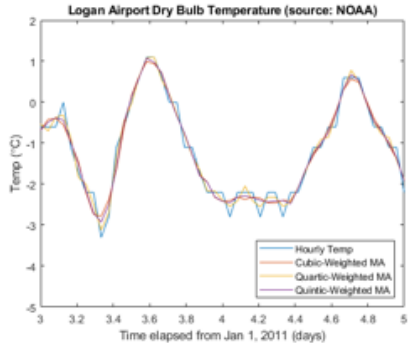
Multivariate Statistics

Machine Learning + Signal Processing

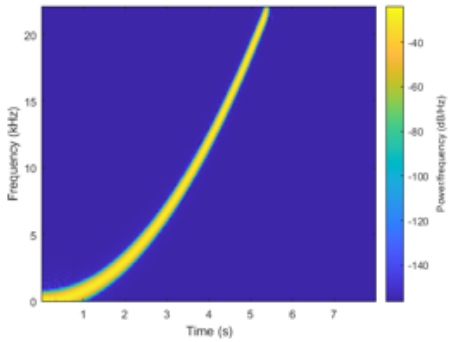
Data Preprocessing



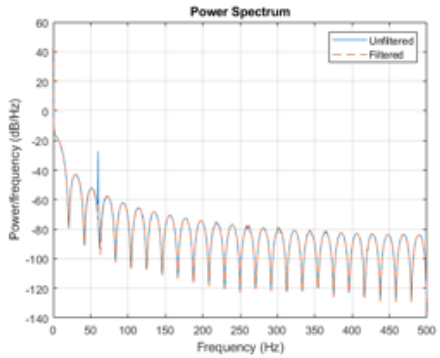
Detrending



Smoothing

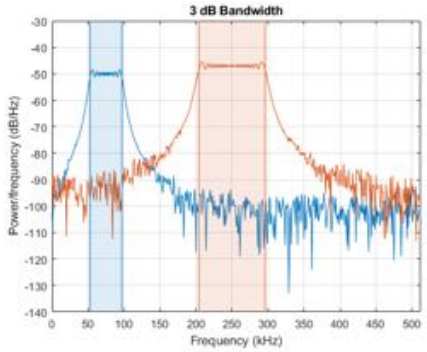


Resampling

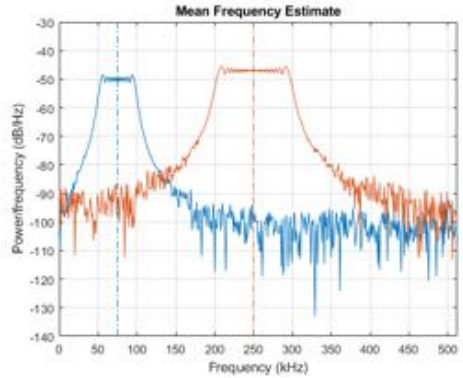


Filtering

Feature Engineering



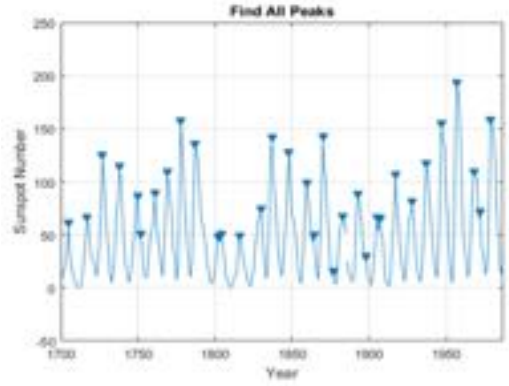
Bandwidth measurements



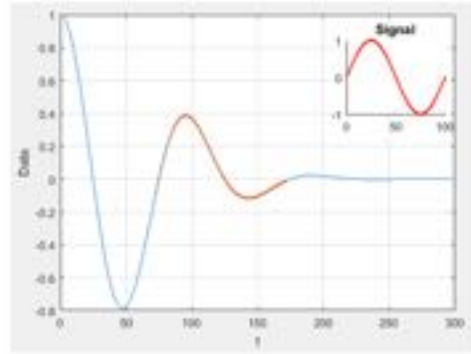
Spectral statistics

Frequency domain

Time domain



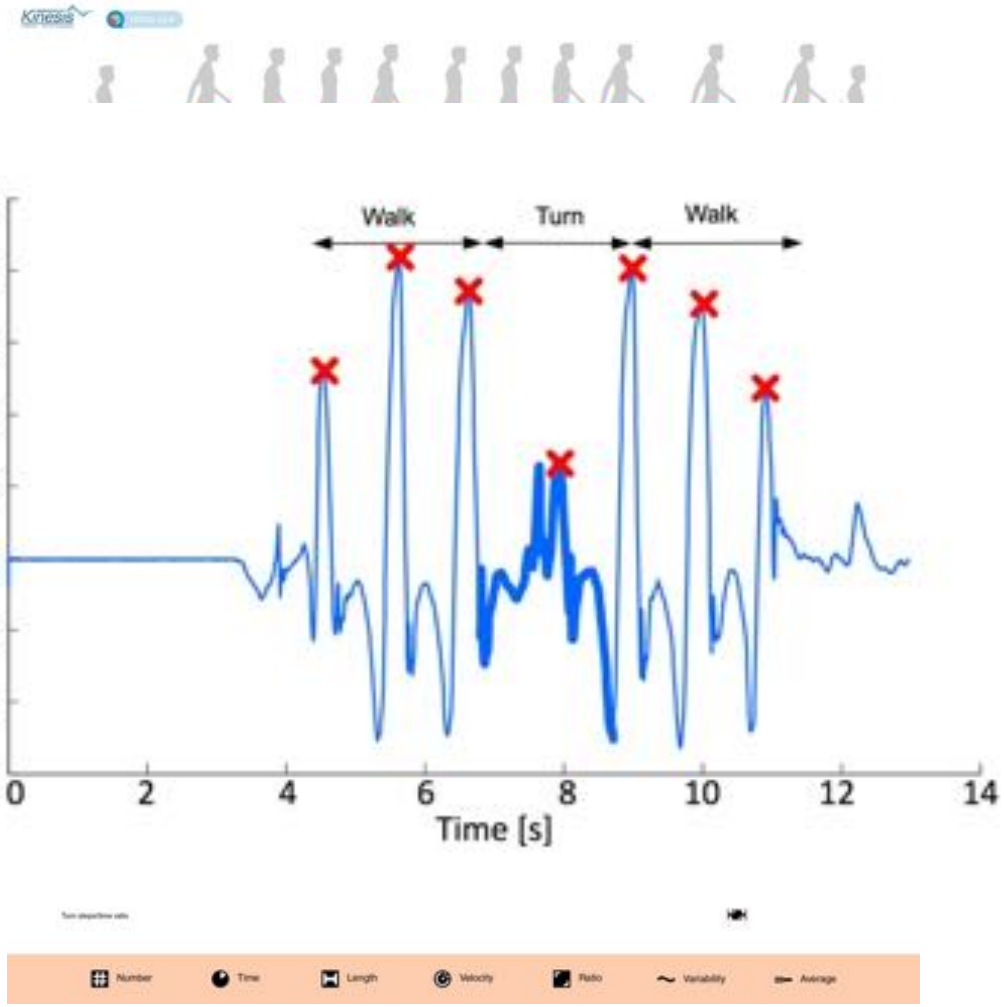
Find peaks



Find signal patterns

Kinesis Health Technologies

Predicting a patient's fall risk with machine learning.



From Desktop to Production

Reasons for Updates:

- Found a better model
- New data became available
- Business needs change
- …

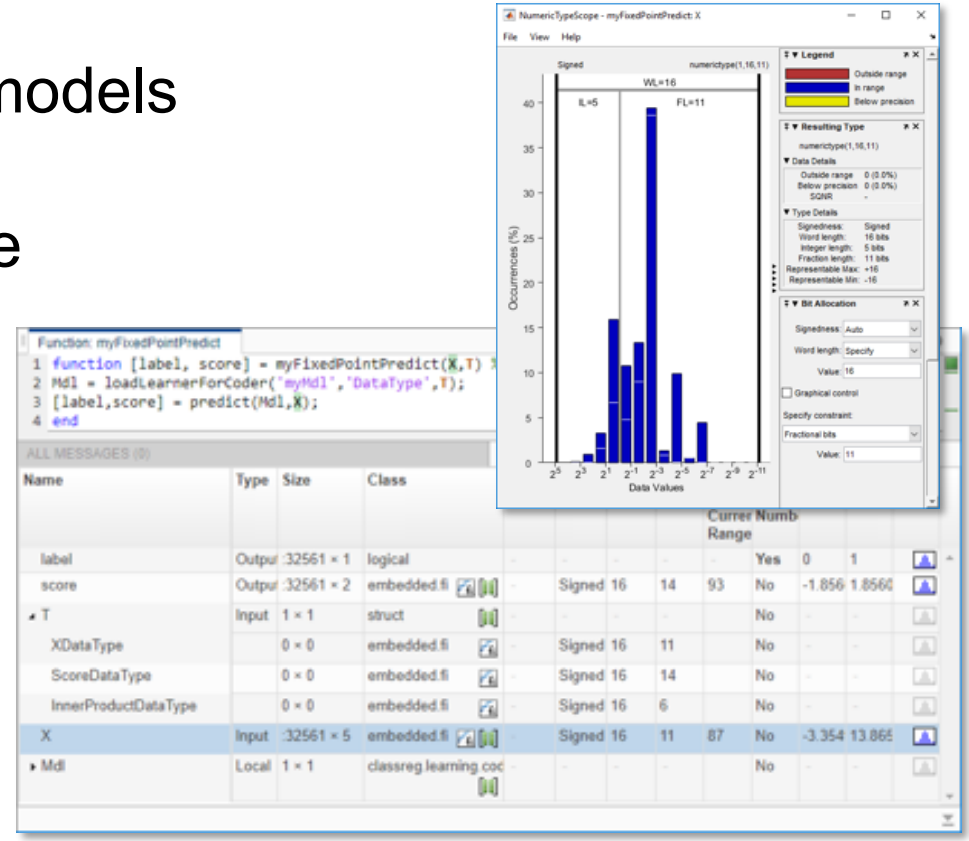
Automatic C/C++ Code Generation

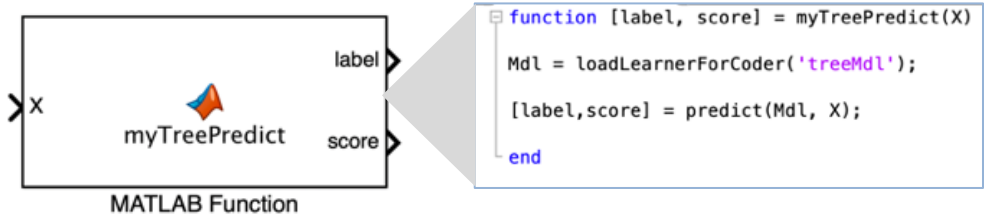
1. Prediction for most Classification and Regression models

2. Update deployed models without regenerating code
   - SVM, Decision Trees, Linear Models

1. Fixed-Point support
   - SVM, Decision Trees, Ensemble of Trees
   - Shallow Neural Network (through Simulink)

1. Integrate with Simulink models as MATLAB Function Block





Integrate MATLAB with Other Languages

Examples of Successful Machine Learning Applications

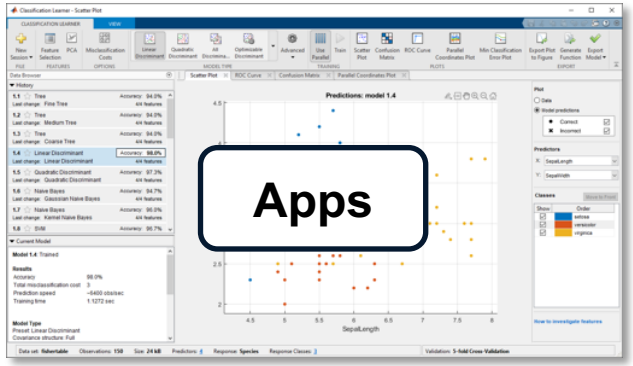
Fleet Data Analytics

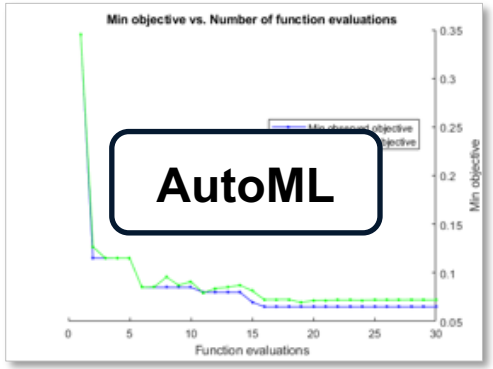
Energy Forecasting

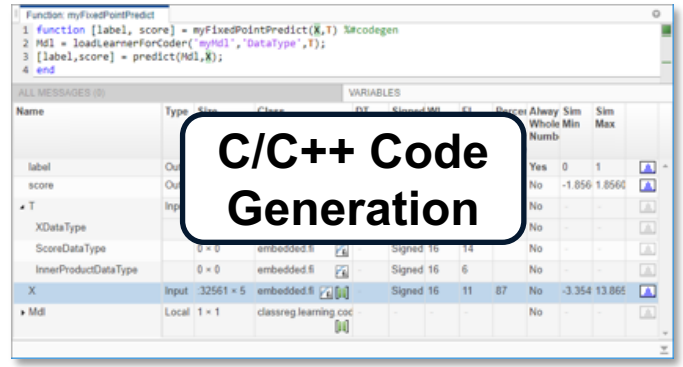
Manufacturing Analytics

New Capabilities

- MATLAB apps
- AutoML
- Signal Processing with Machine Learning
- C/C++ Code Generation







Machine Learning

+

Fleet Data Analytics

Signal Processing

Industry Knowledge

X

Energy Forecasting

Manufacturing Analytics

Application Knowledge

Medical Devices

Mining

Learn More

Training Courses

MATLAB Fundamentals (3 days)

MATLAB for Data Processing and Visualization (1 day)

Processing Big Data with MATLAB (1 day)

Statistical Methods in MATLAB (2 days)

Machine Learning with MATLAB (2 days)

Signal Preprocessing and Feature Extraction with MATLAB (1 day)

Deep Learning with MATLAB (2 days)

Accelerating and Parallelizing MATLAB Code (2 days)

- Exploratory Data Analysis
- Data Processing and Feature Engineering
- Predictive Modeling and Machine Learning
- Data Science Project

MATLAB EXPO

MathWorks