Data Analytics and Machine Learning with MATLAB

Mandar Gujrathi
Applications Engineer
Mandar.Gujrathi@mathworks.com.au

Aeronautics

Railway Systems

Automotive

Retail

Finance

Off-highway vehicles

Condition Monitoring

Prognostics

Retail Analytics

Fleet Analytics

Data Analytics

Operational Analytics

Process Analytics

Risk Analysis

Health Monitoring

Supply Chain

Mfg Process Analytics

Asset Analytics

Healthcare Analytics

Internet

Industrial Automation

Logistics

Oil & Gas

Clean Energy

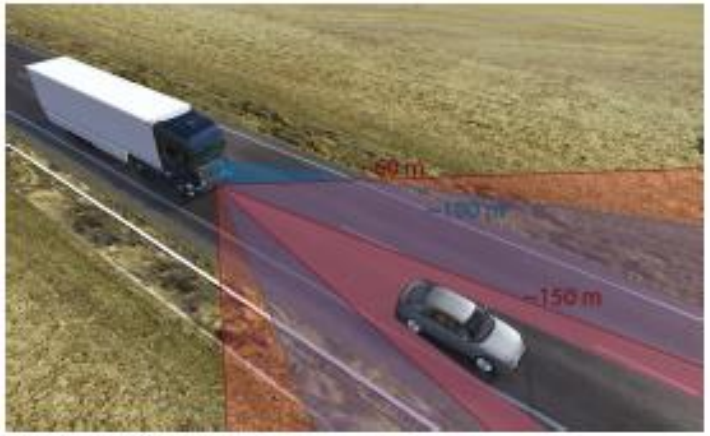
Medical Devices

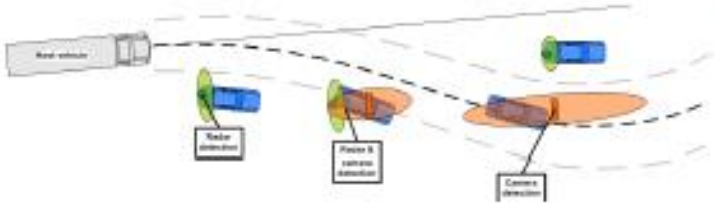
Healthcare Management

Automatic Emergency Braking using Sensor Fusion & Analytics

Sensor fusion

Two sensors -> One "truth"





Sensors have different advantages

- Radar
  - \+ Range (longitudinal)
  - \+ Relative velocity
  - \+ Solid object reflection
  - \- No shapes
  - \- Lateral position
- Camera
  - \+ Object type
  - \+ Object width
  - \+ Lateral position
  - \- Range
  - \- Optical illusions

2015-08-24 Jonny Andersson

SCANIA

Automatic Emergency Braking using Sensor Fusion & Analytics

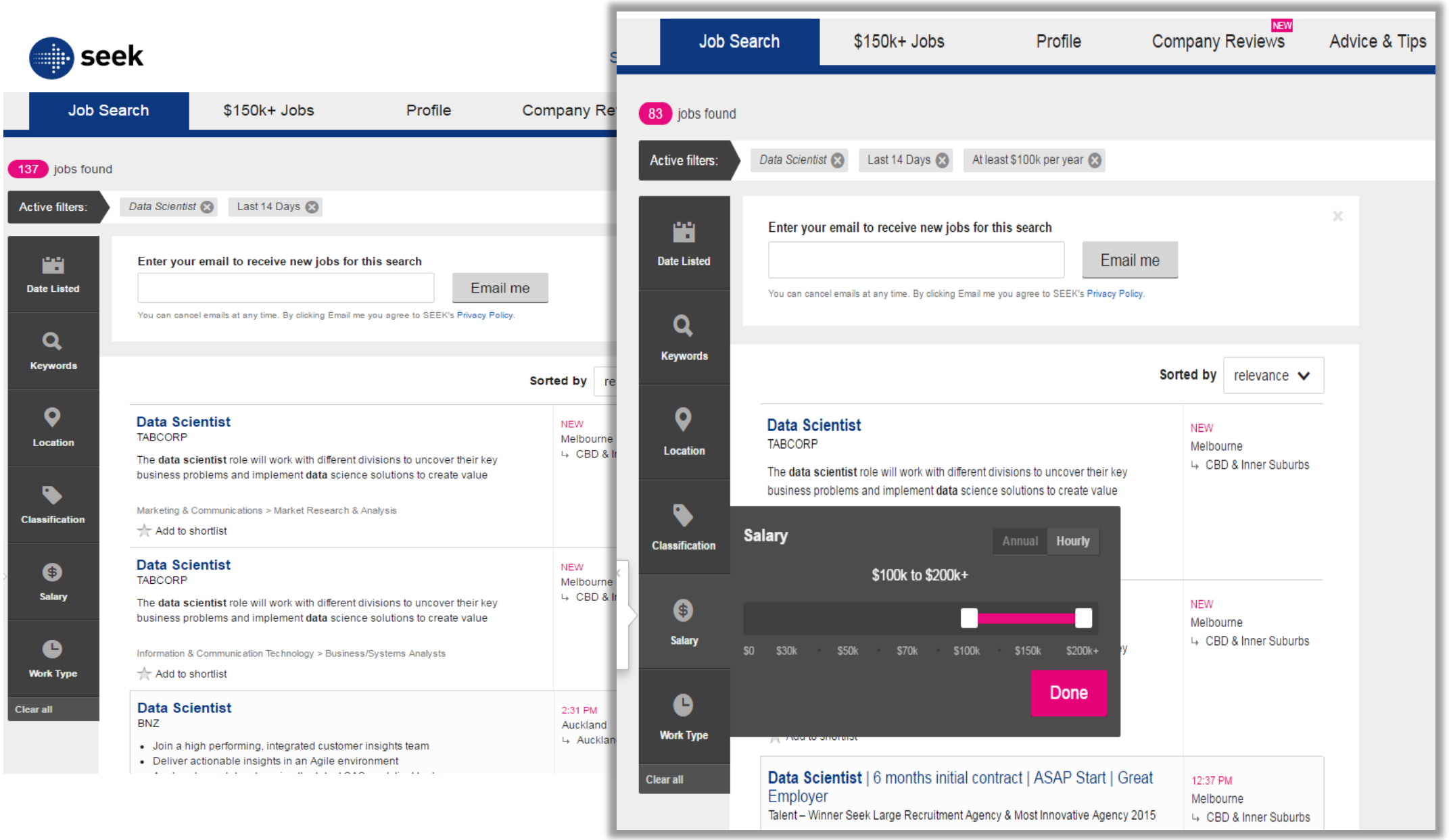
Traits of Data Analytics applications

1. **How can I handle diverse and/or Big Data ?**

2. **How to handle/create advanced algorithms, e.g., Machine Learning ?**

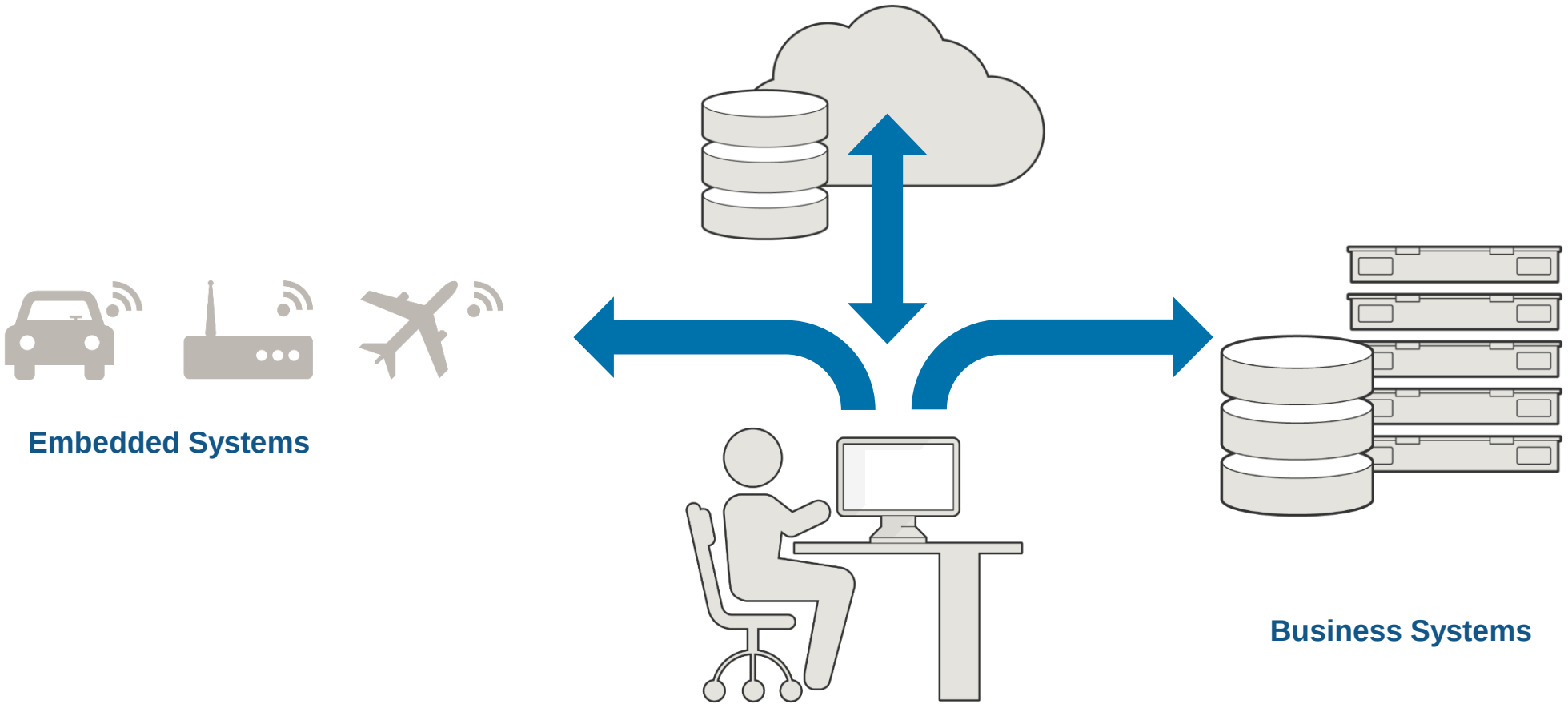
3. **How can I integrate these algorithms?**

Market Demands



The *role* of a Data Scientist

DATA
- **Engineering, Scientific, and Field**
- **Business and Transactional**



Embedded Systems

Business Systems

Researchers/ Data Scientist/ Students

Why MATLAB?

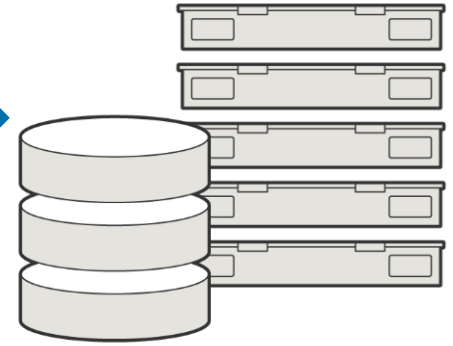
1. Analytics that increasingly require **both business and engineering data**

DATA
- Engineering, Scientific, and Field
- Business and Transactional

3. Develop **embedded systems** with analytics powered functionality

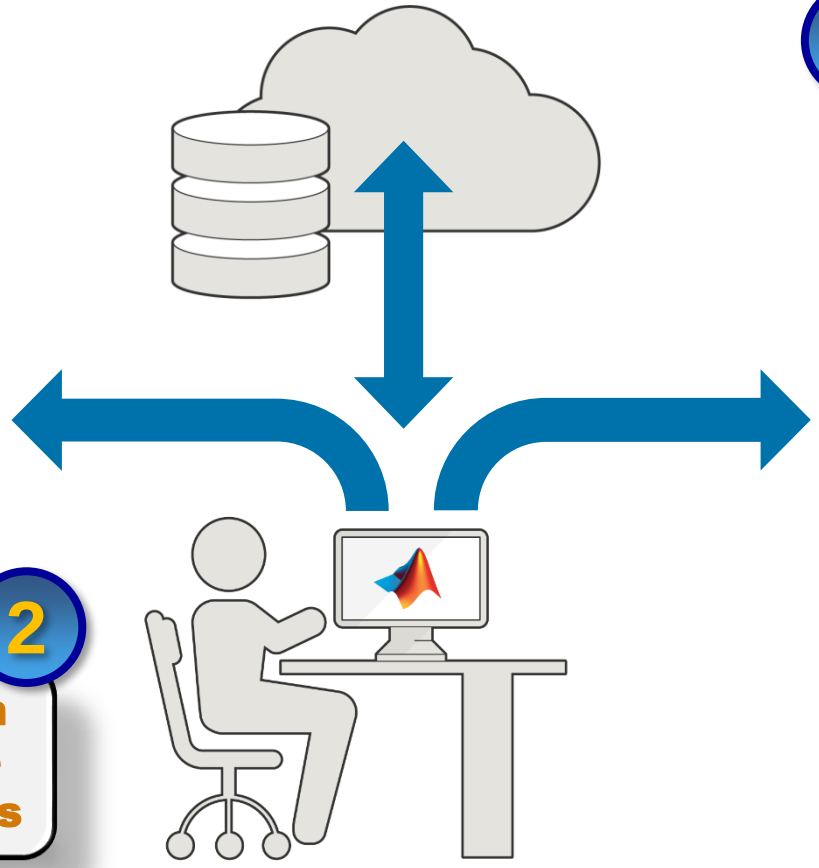
Embedded Systems

4. Develop analytics to run on **both enterprise and embedded platforms**



Business Systems

2. Enable **Domain Experts to be Data Scientists**



Researchers/ Data Scientist/ Students

Why MATLAB?

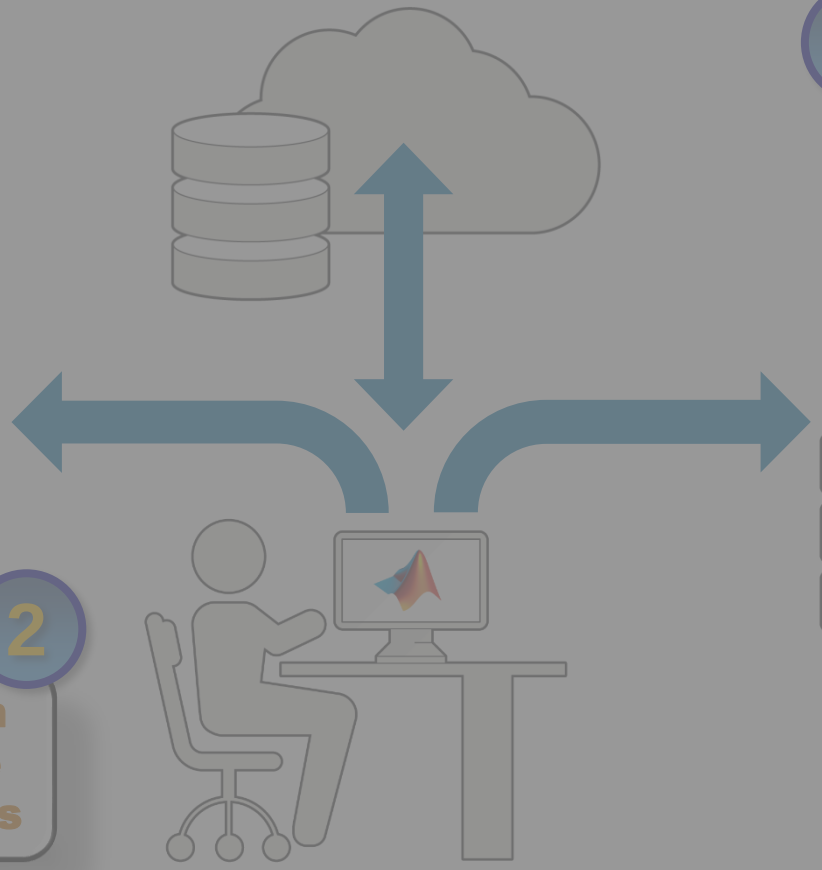
1 Analytics that increasingly require **both business and engineering data**

DATA
- Engineering, Scientific, and Field
- Business and Transactional

3 Develop **embedded systems** with analytics powered functionality

4 Develop analytics to run on **both enterprise and embedded platforms**

Smarter Embedded Systems



2 Enable **Domain Experts to be Data Scientists**

Data Scientist

Business Systems

Business and/or Engineering, MATLAB is for all

Business and Transactional Data

Repositories
- Databases
- **Hadoop**

File I/O
- Text
- Spreadsheet
- XML

Web Sources
- HTML
- Mapping
- Financial datafeeds
- **RESTful**
- **JSON**

Engineering, Scientific, and Field Data

File I/O
- Text
- Spreadsheet
- XML
- CDF/HDF
- Image
- Audio
- Video
- Geospatial

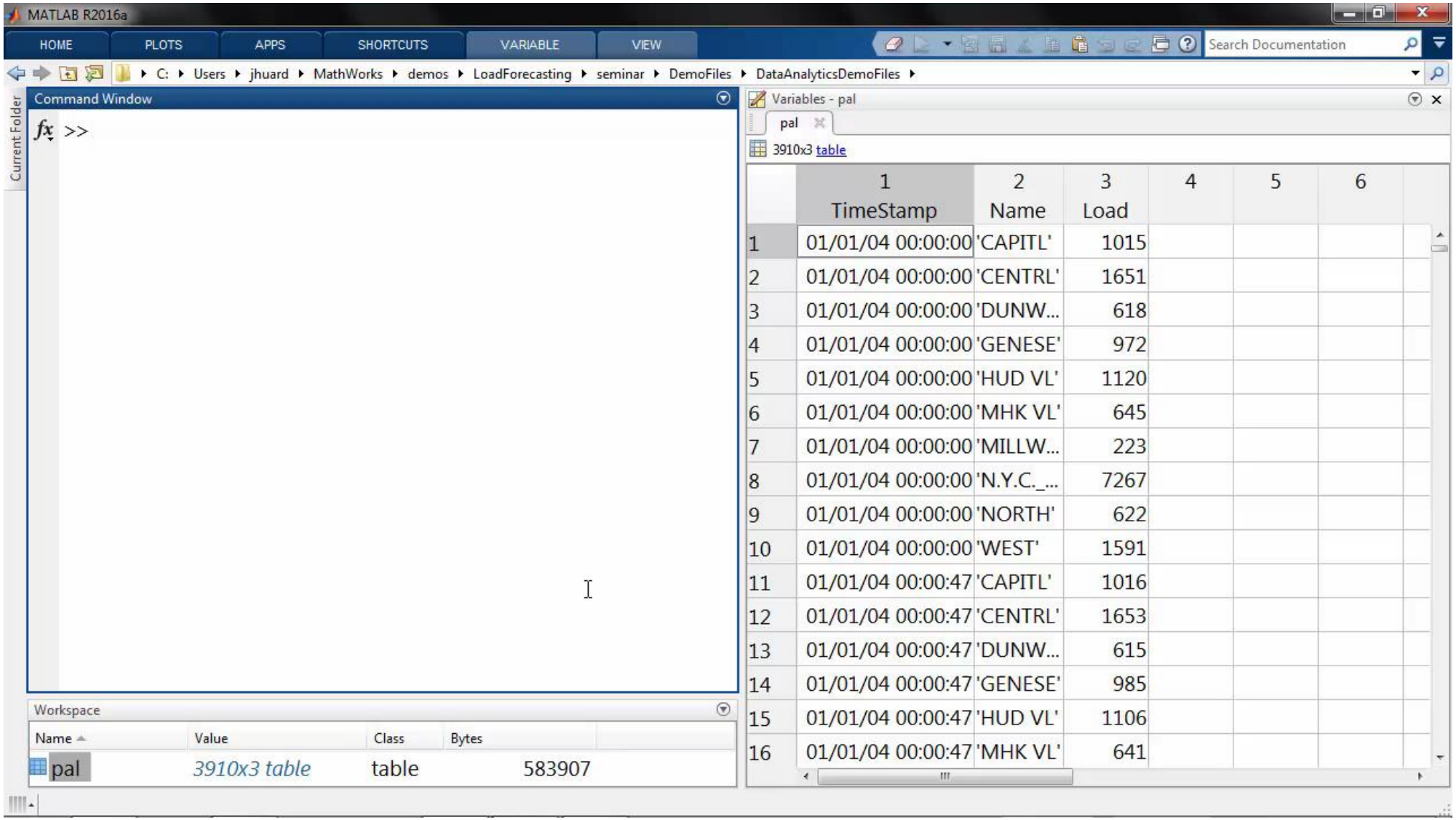
Communication Protocols
- CAN (Controller Area Network)
- DDS (Data Distribution Service)
- OPC (OLE for Process Control)
- XCP (eXplicit Control Protocol)

Real-Time Sources
- Sensors
- GPS
- Instrumentation
- Cameras
- Communication systems
- Machines (embedded systems)

"No matter what industry our client is in, and no matter what data they ask us to analyze—text, audio, images, or video—MATLAB enables us to provide clear results faster."

Dr. G Subrahmanya VRK Rao, Cognizant

Data handling and visualization



High-quality domain-specific libraries that just work

Data type	Common Techniques for Deriving Features
Sensor data	Signal Processing
Image and video data	Image Processing Computer Vision
Transactional data	Statistics & Machine Learning

Why MATLAB?

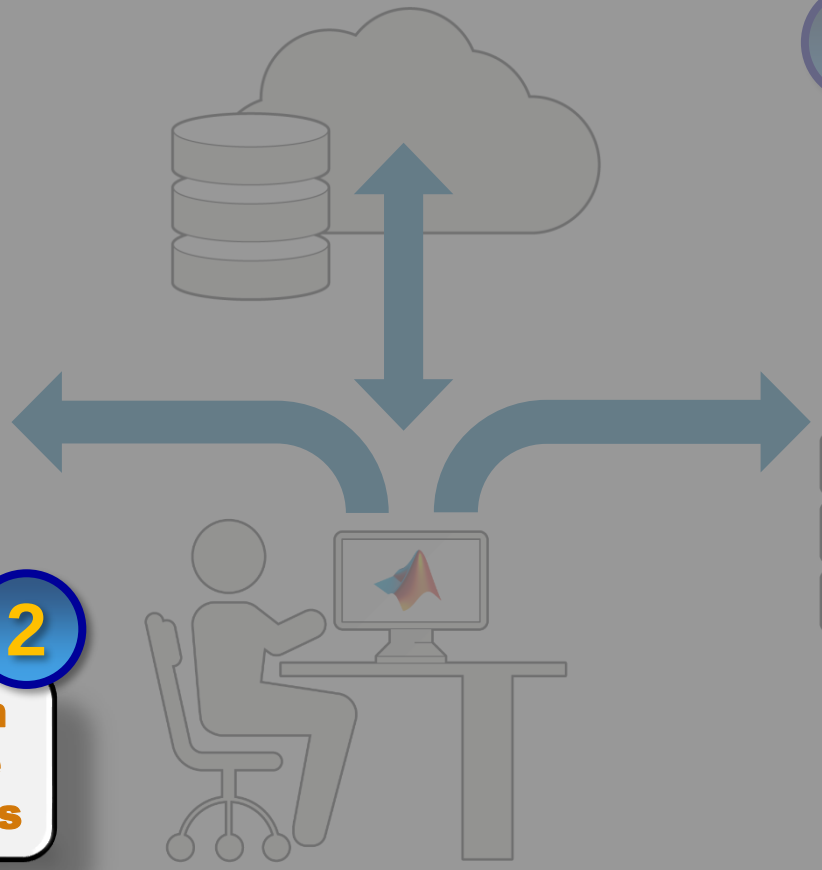
1. Analytics that increasingly require **both business and engineering data**

DATA
- Engineering, Scientific, and Field
- Business and Transactional

3. Develop **embedded systems** with analytics powered functionality

4. Develop analytics to run on **both enterprise and embedded platforms**

Smarter Embedded Systems

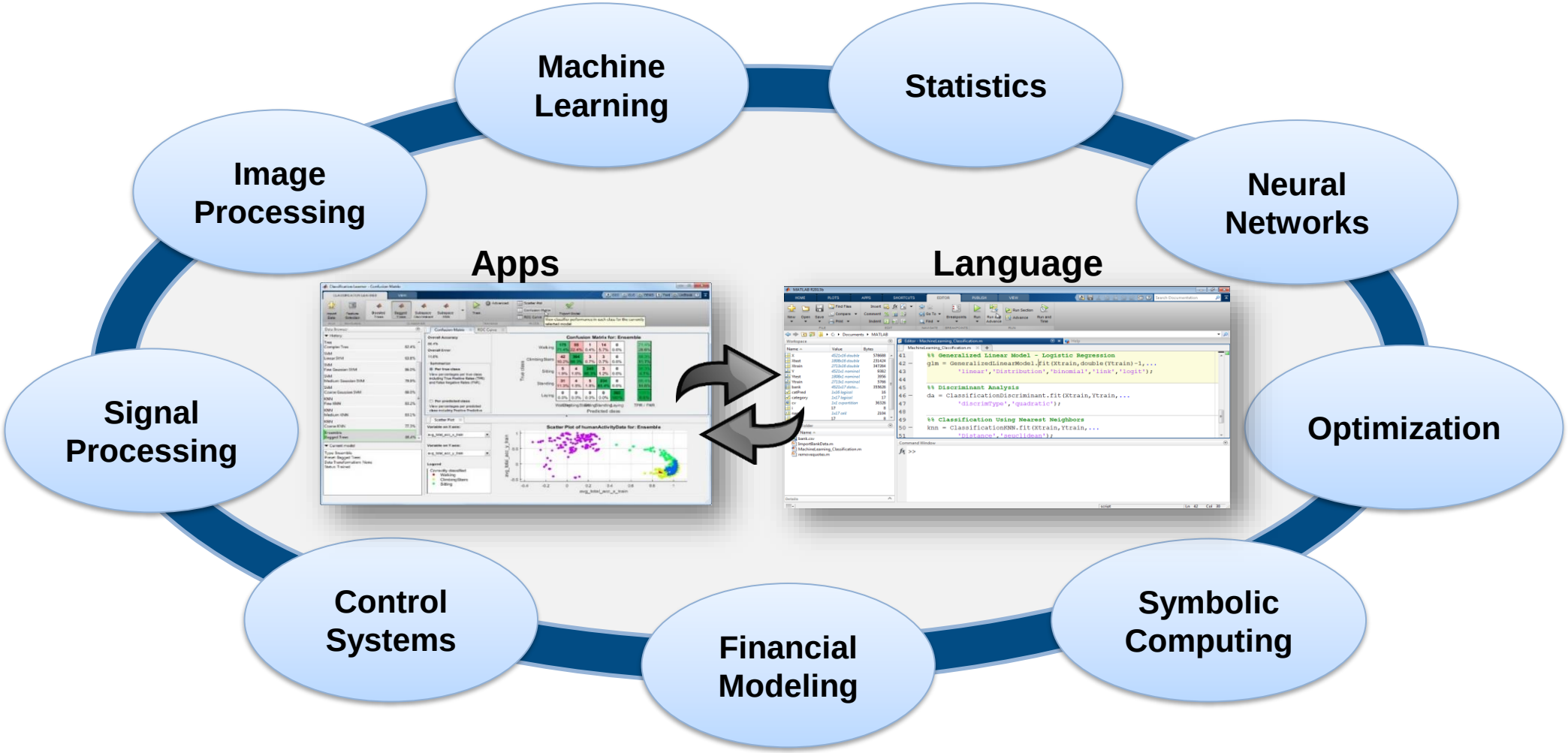


Data Scientist

Business Systems

2. Enable **Domain Experts to be Data Scientists**

Enabling Domain Experts to be Data Scientists



Built-in algorithms

Clustering

Hierarchical Clustering
Produce nested sets of clusters

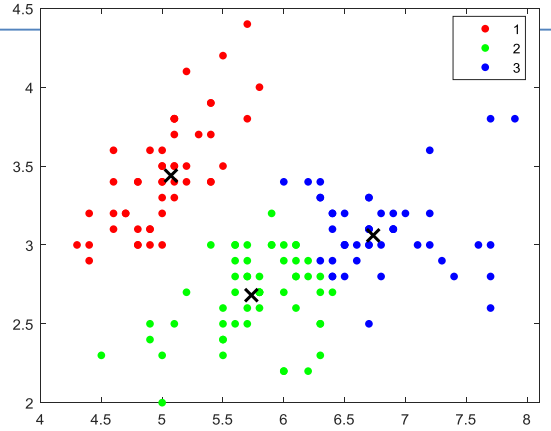
k-Means and k-Medoids Clustering
Cluster by minimizing mean or medoid distance, calculate Mahalan

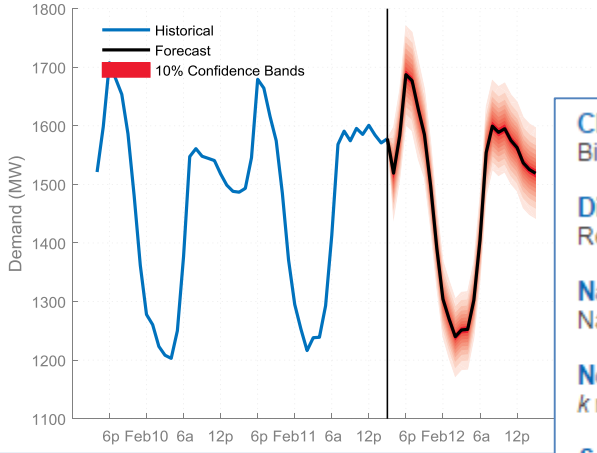
Gaussian Mixture Models
Cluster based on Gaussian mixture models using the EM algorithm

Nearest Neighbors
Find nearest neighbors using exhaustive search or *kd*-tree s

Hidden Markov Models
Markov models for data generation

Cluster Visualization and Evaluation
Plot clusters of data and evaluate optimal number of clusters





Regression

Linear Regression
Multiple, stepwise, multivariate regression models, and mo

Generalized Linear Models
Logistic regression, multinomial regression, Poisson regre

Nonlinear Regression
Nonlinear fixed- and mixed-effects regression models

Support Vector Machine Regression
Support vector machines for regression models

Gaussian Process Regression
Gaussian process regression models (kriging)

Regression Trees
Binary decision trees for regression

Regression Tree Ensembles
Random forests, boosted and bagged regression trees

Classification

Classification Trees
Binary decision trees for multiclass learning

Discriminant Analysis
Regularized linear and quadratic discriminant analysis

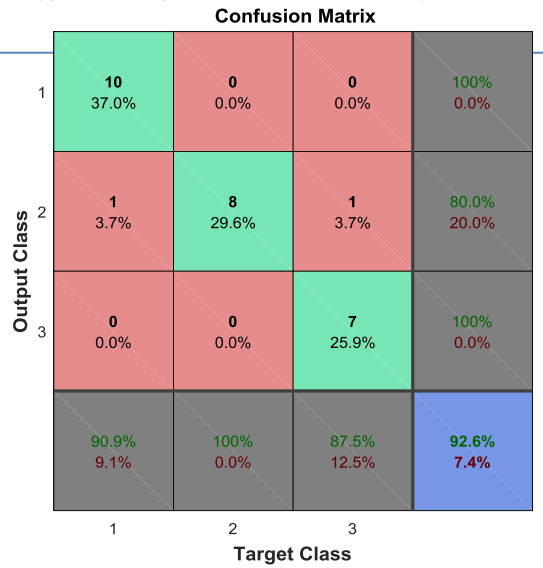
Naive Bayes
Naive Bayes model with Gaussian, multinomial, or kernel predictors

Nearest Neighbors
k nearest neighbors classification using *Kd*-tree search

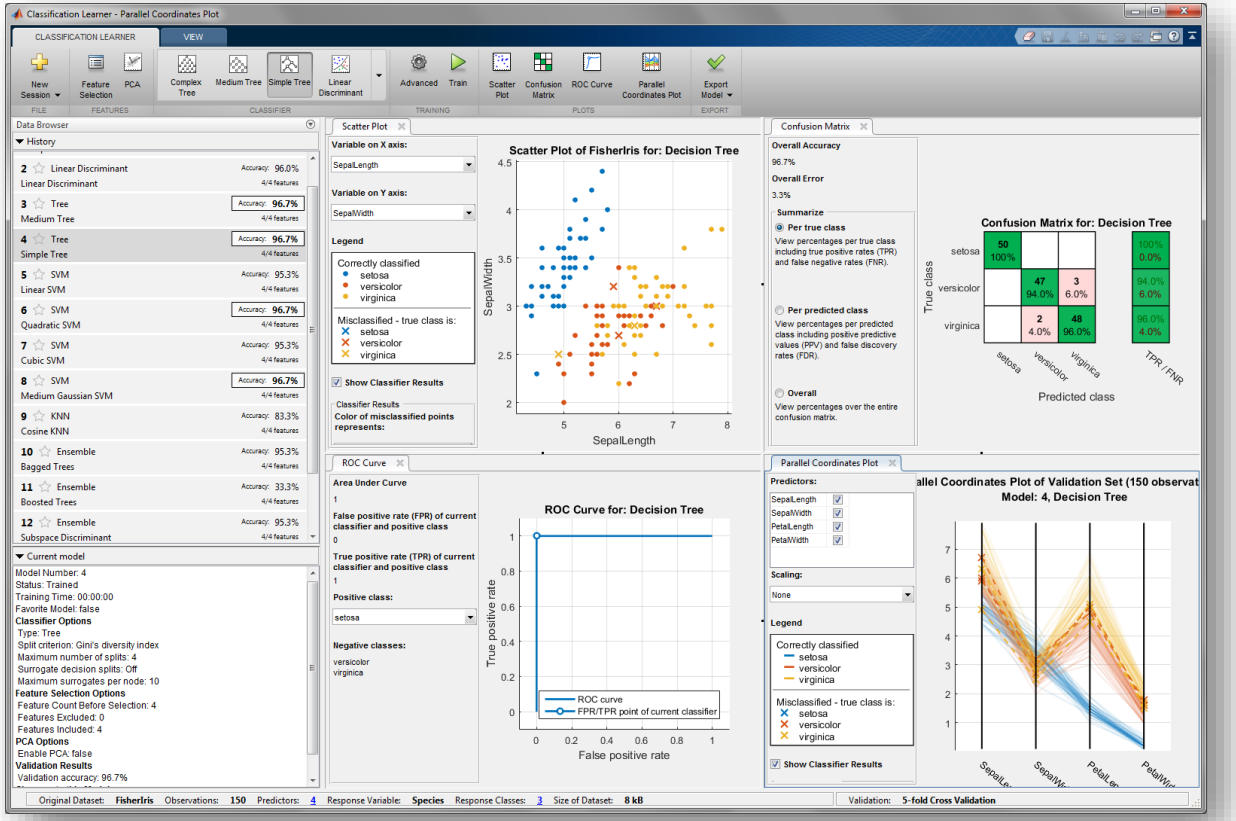
Support Vector Machine Classification
Support vector machines for binary or multiclass classification

Classification Ensembles
Boosting, random forest, bagging, random subspace, and ECOC ensembles for multiclass learning

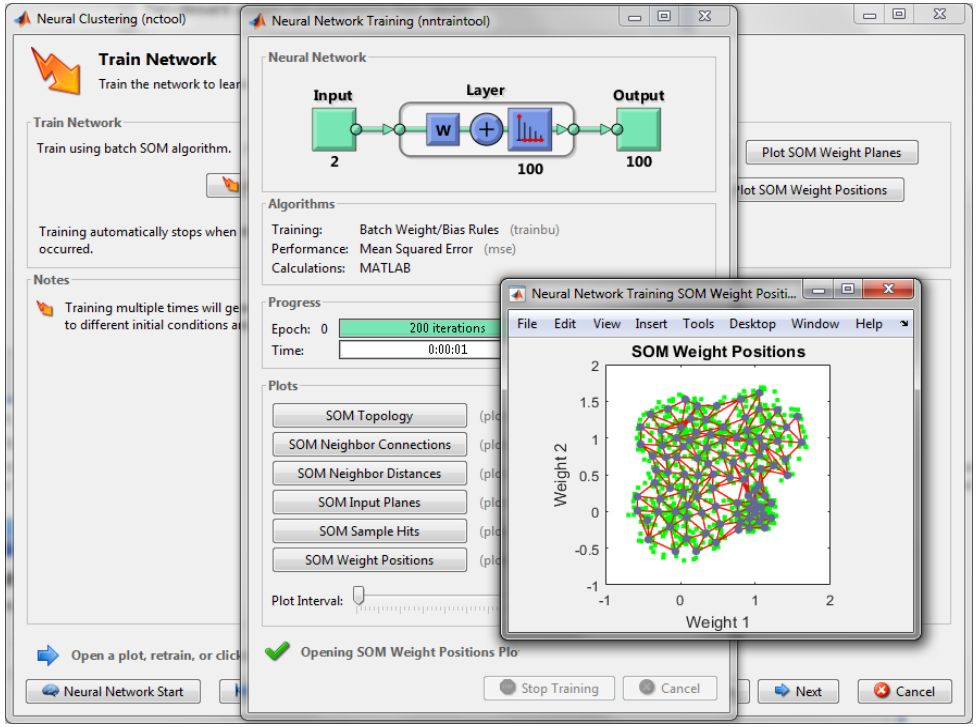
Model Building and Assessment
Feature selection, cross validation, predictive performance evaluation, classification accuracy comparison tests



Interactive Apps to focus on machine learning, not programing



Classification Learner App



Neural network Apps

Features

- Train models
- Assess results
- Export models to the MATLAB or generate MATLAB code

Data Analytics for predictive maintenance

- Example: faulty braking system leads to windmill disaster
  - https://youtu.be/-YJuFvjtM0s?t=39s
- Wind turbines cost millions of dollars
- Failures can be dangerous
- Maintenance also very expensive and dangerous

Types of Maintenance

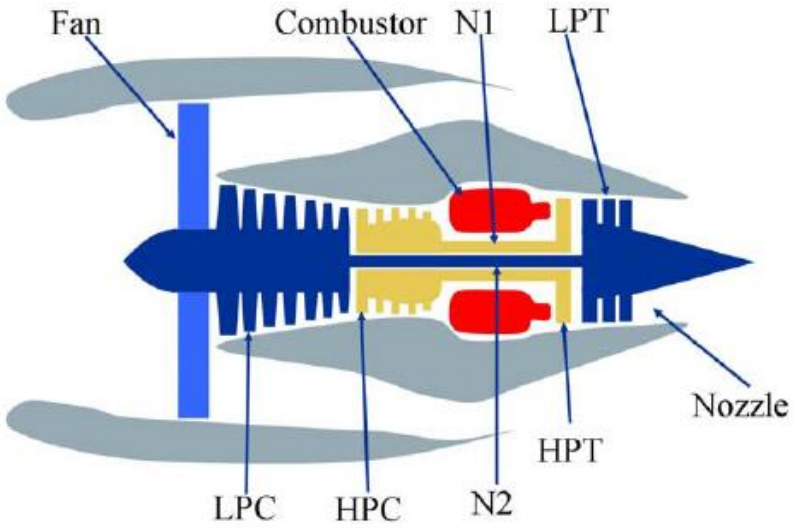
- Reactive – Do maintenance once there's a problem
  - Example: replace car battery when it has a problem
  - Problem: unexpected failures can be expensive and potentially dangerous
- Scheduled – Do maintenance at a regular rate
  - Example: change car's oil every 5,000 miles
  - Problem: unnecessary maintenance can be wasteful; may not eliminate all failures
- Predictive – Forecast when problems will arise
  - Example: certain GM car models forecast problems with the battery, fuel pump, and starter motor
  - Problem: difficult to make accurate forecasts for complex equipment

Predictive Maintenance of Turbofan Engine

Sensor data from 100 engines of the same model

Predict and fix failures before they arise

- Import and analyze historical sensor data
- Train model to predict when failures will occur
- Deploy model to run on live sensor data
- Predict failures in real time



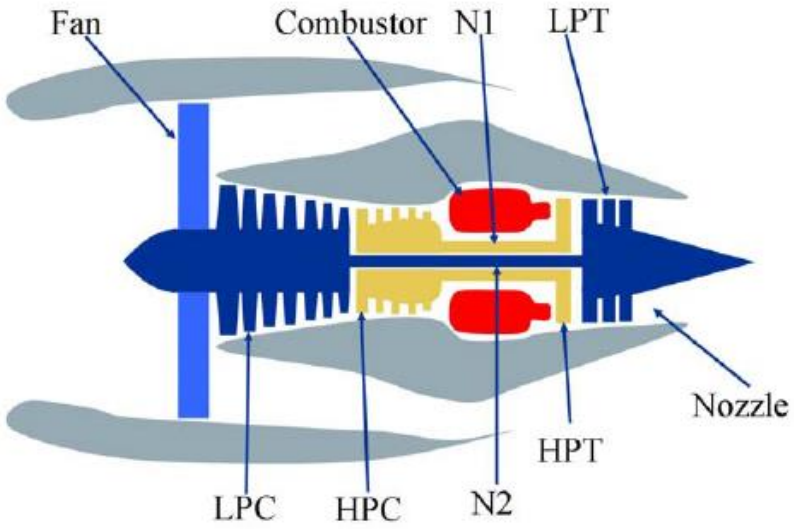
Data provided by NASA PCoE
http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/

Predictive Maintenance of Turbofan Engine

Sensor data from 100 engines of the same model

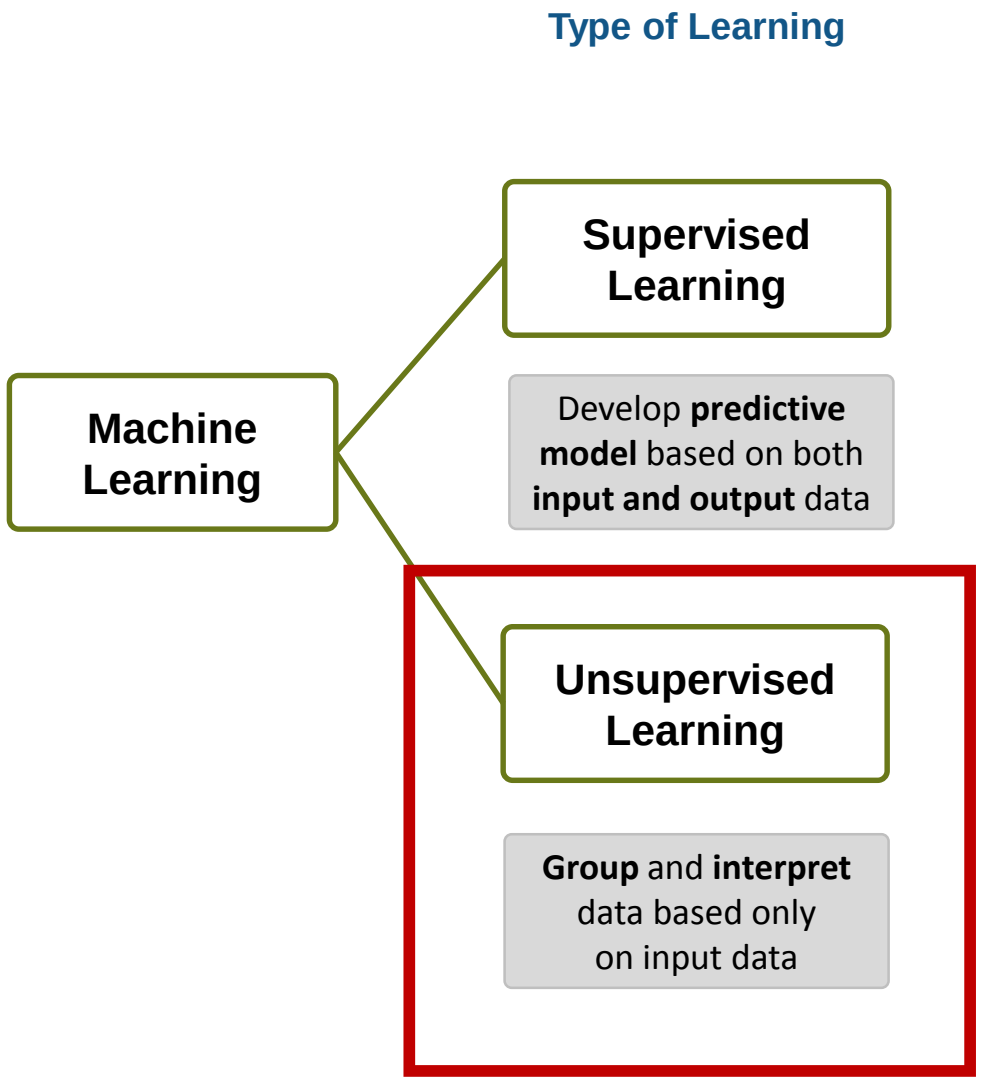
Scenario 1: No data from failures

- Performing scheduled maintenance
- No failures have occurred
- Maintenance crews tell us most engines could run for longer
- Can we be smarter about how to schedule maintenance **without** knowing what failure looks like?

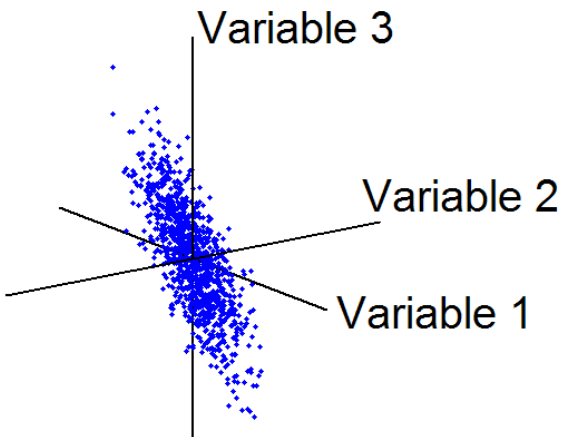


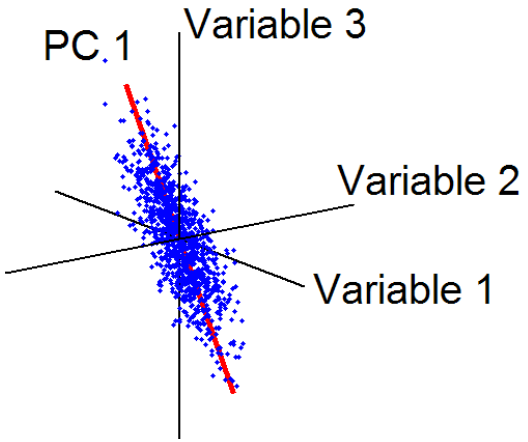
Data provided by NASA PCoE
http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/

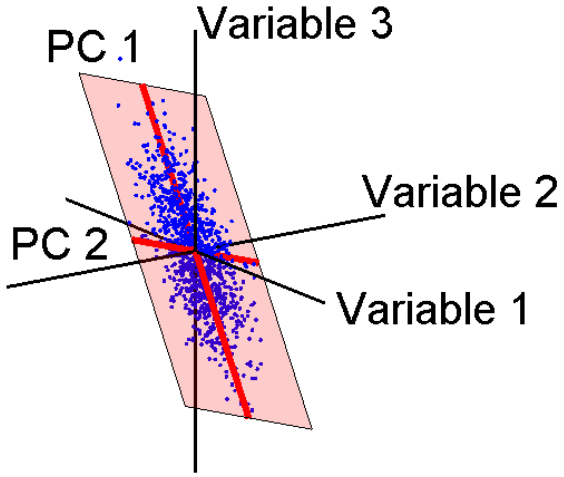
Overview – Machine Learning

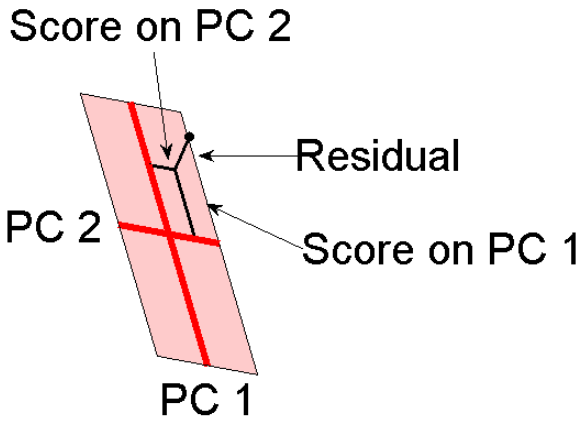


Principal Components Analysis – what is it doing?

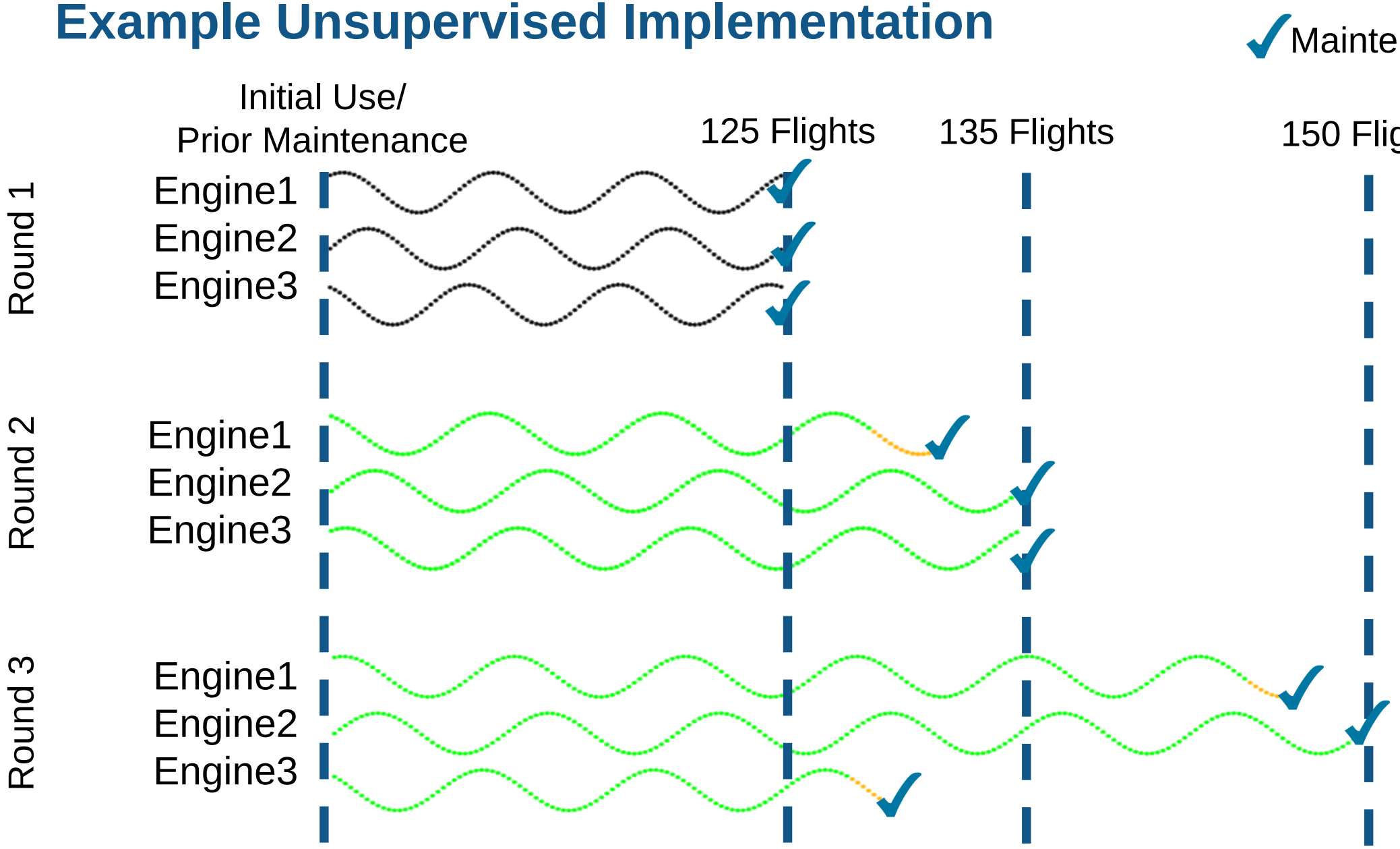








Example Unsupervised Implementation

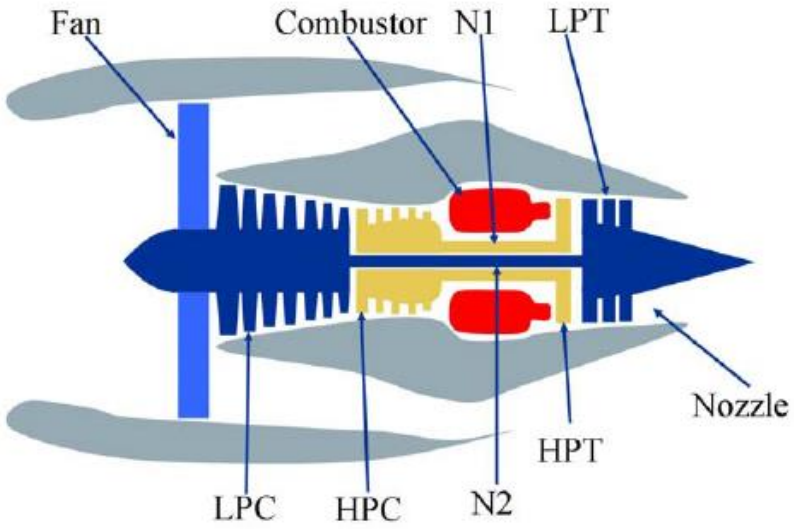


Predictive Maintenance of Turbofan Engine

Sensor data from 100 engines of the same model

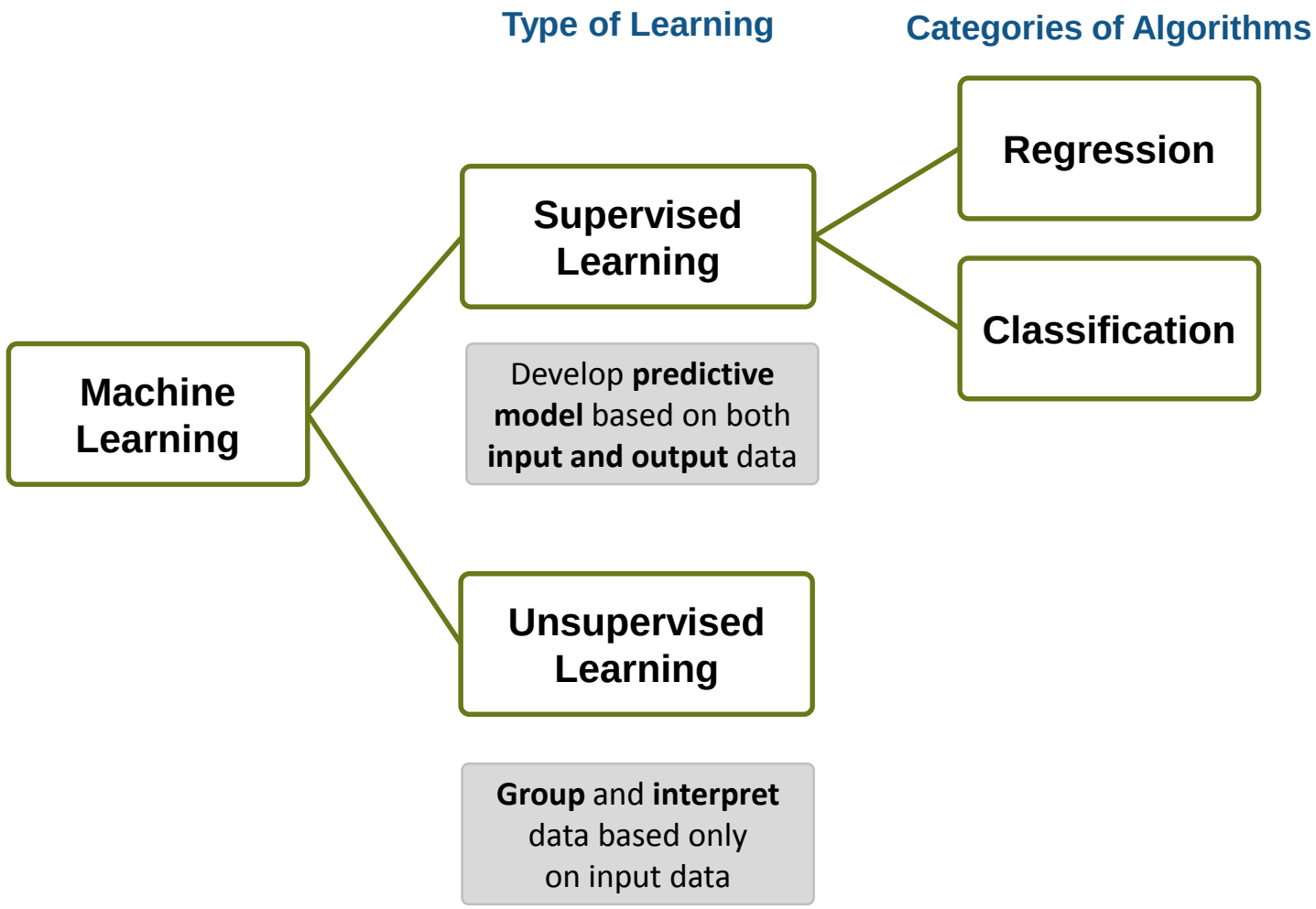
Scenario 2: Have failure data

- Performing scheduled maintenance
- Failures still occurring (maybe by design)
- Search records for when failures occurred and gather data preceding the failure events
- Can we predict how long until failures will occur?

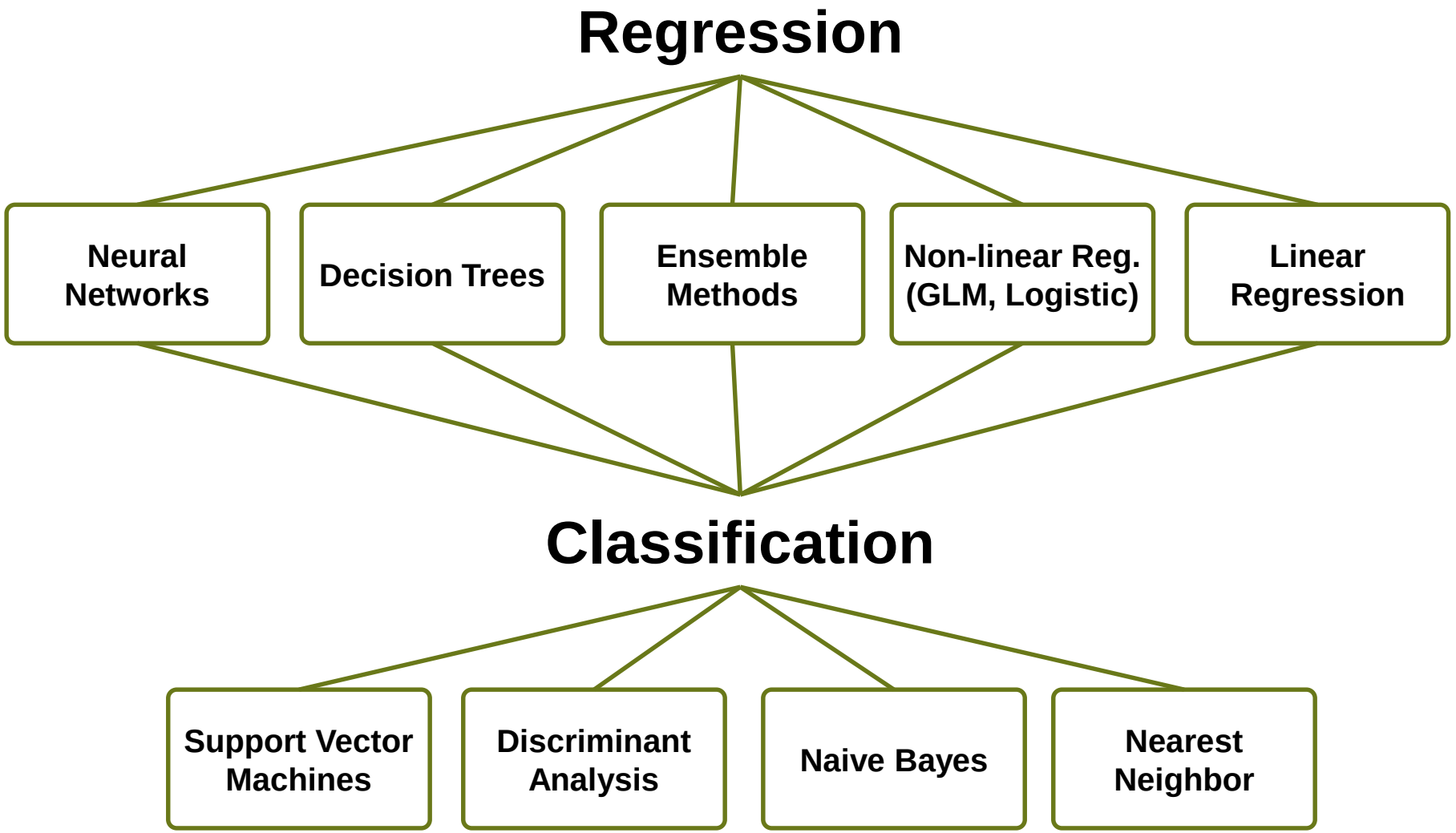


Data provided by NASA PCoE
http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/

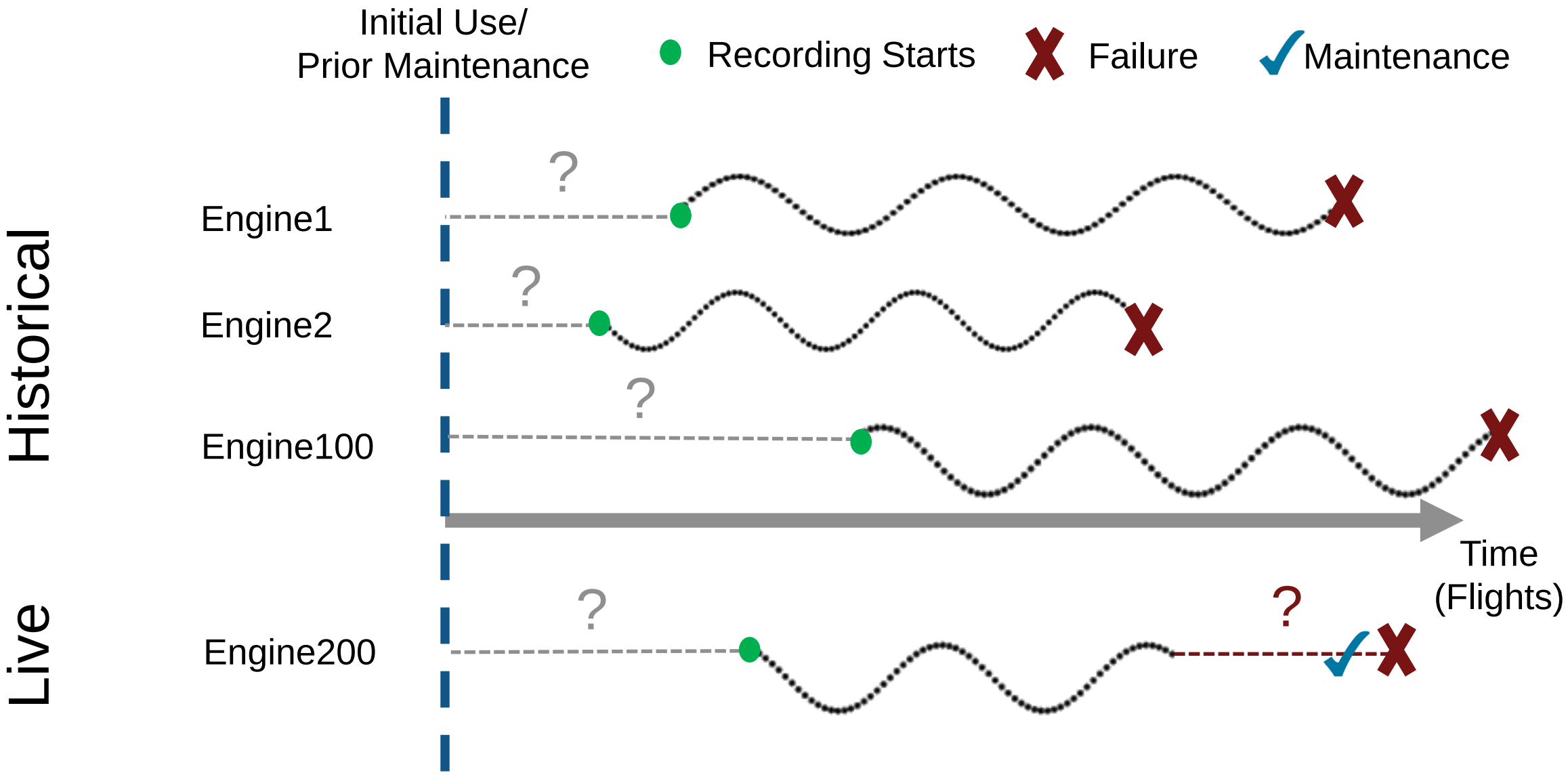
Overview – Machine Learning



Supervised Learning: Functionalities that speak MATH



How Data was Recorded



Why MATLAB?

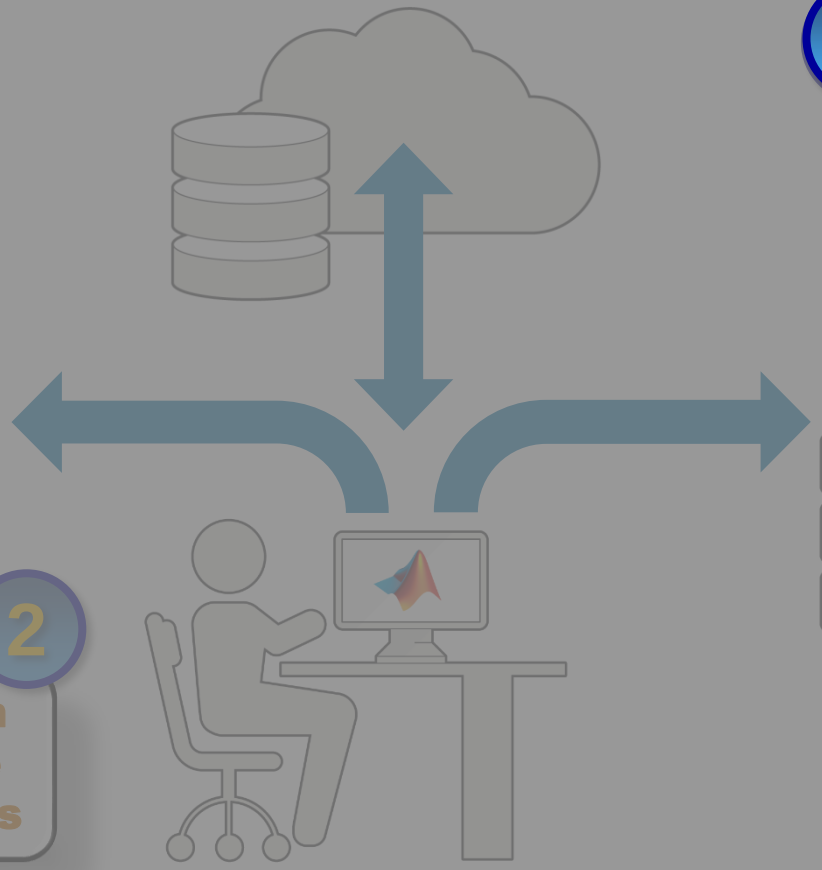
1 Analytics that increasingly require **both business and engineering data**

DATA
- Engineering, Scientific, and Field
- Business and Transactional

3 Develop **embedded systems** with analytics powered functionality

4 Develop analytics to run on **both enterprise and embedded platforms**

Smarter Embedded Systems



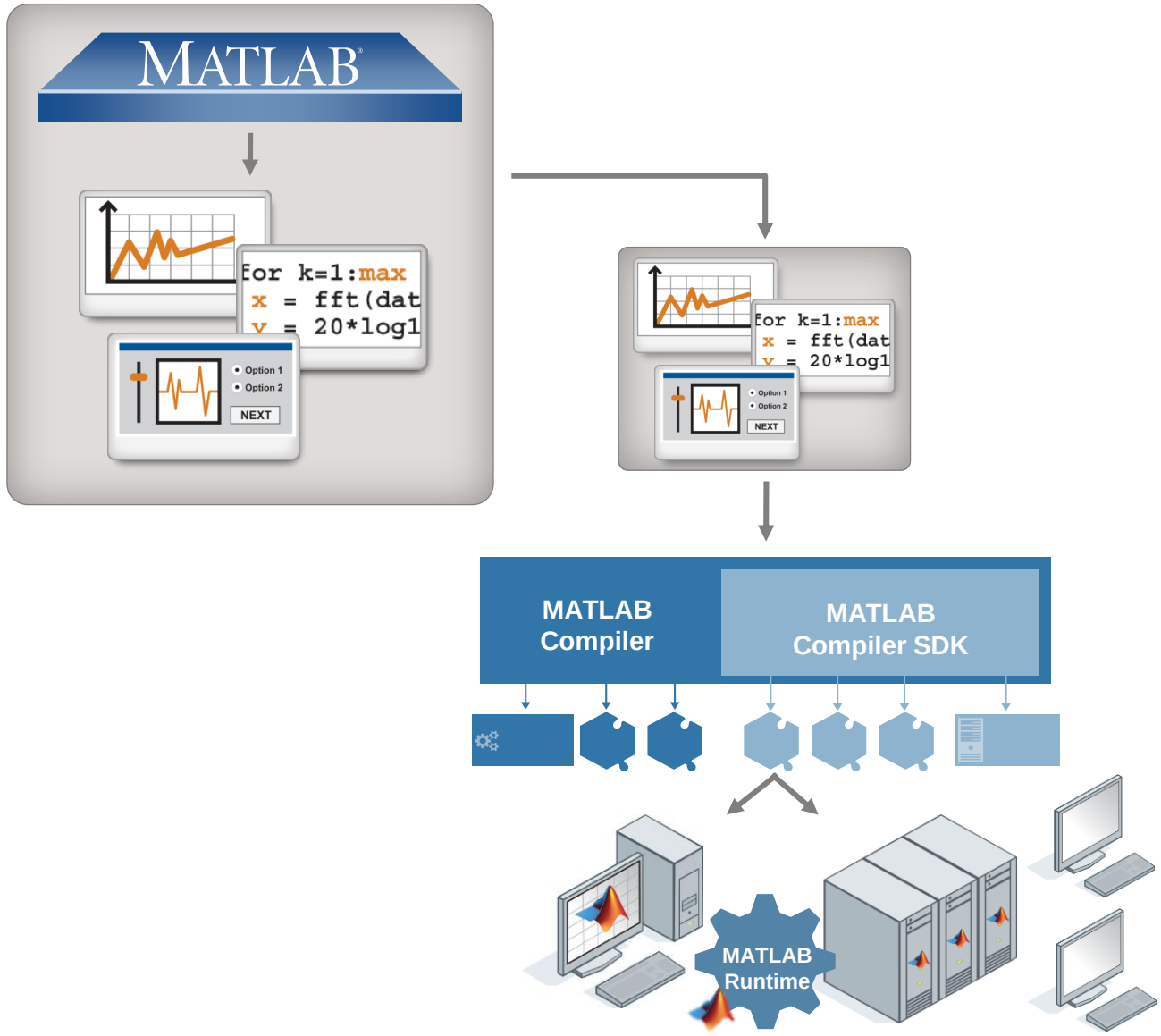
2 Enable **Domain Experts to be Data Scientists**

Data Scientist

Business Systems

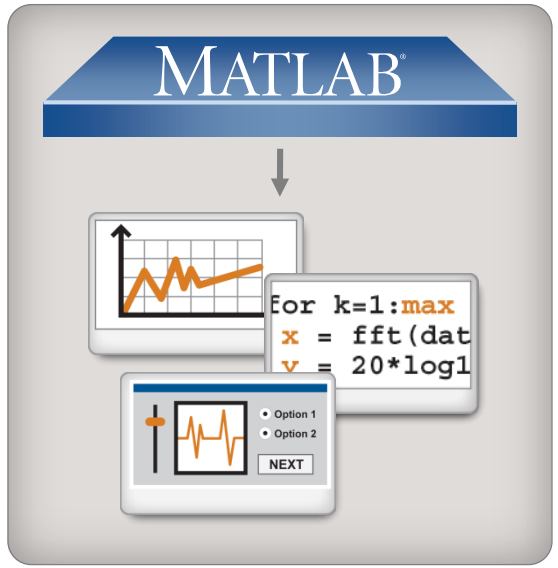
Integrate analytics with your enterprise systems

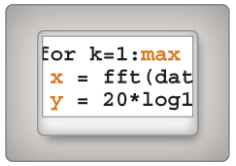
MATLAB Compiler



Integrate analytics with your enterprise systems

MATLAB Coder





MATLAB Coder

.c,.cpp

.exe .lib .dll

Airbus
Battery management

GM
Climate control

Festo
Industrial robots

Sonova
Hearing implants

Weinmann
Transport ventilator

ABB
Smart Grid controller

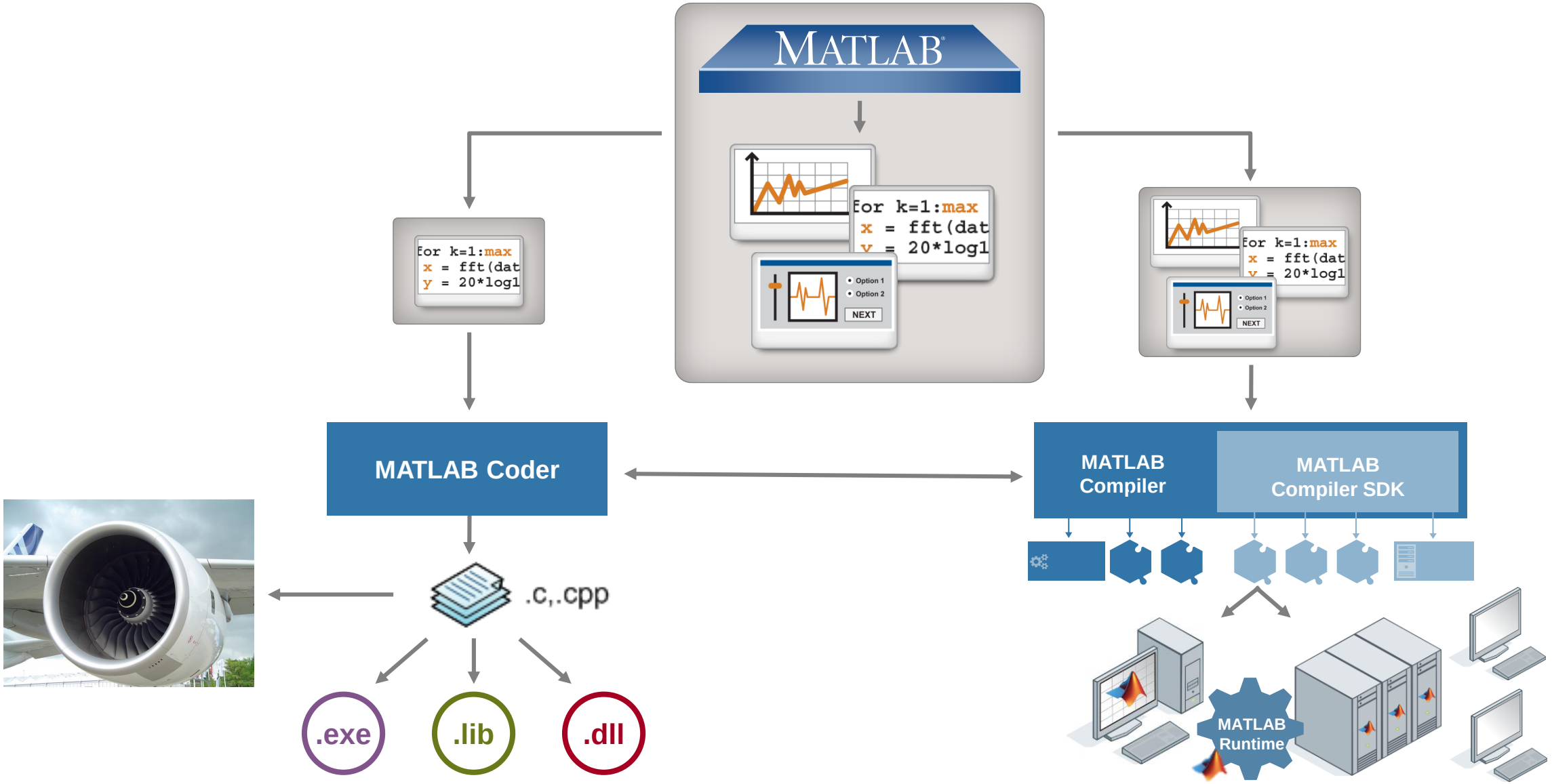
manroland
Printing presses

FLIR
Thermal imaging

Daimler
Cruise controller

Integrate analytics with your enterprise systems

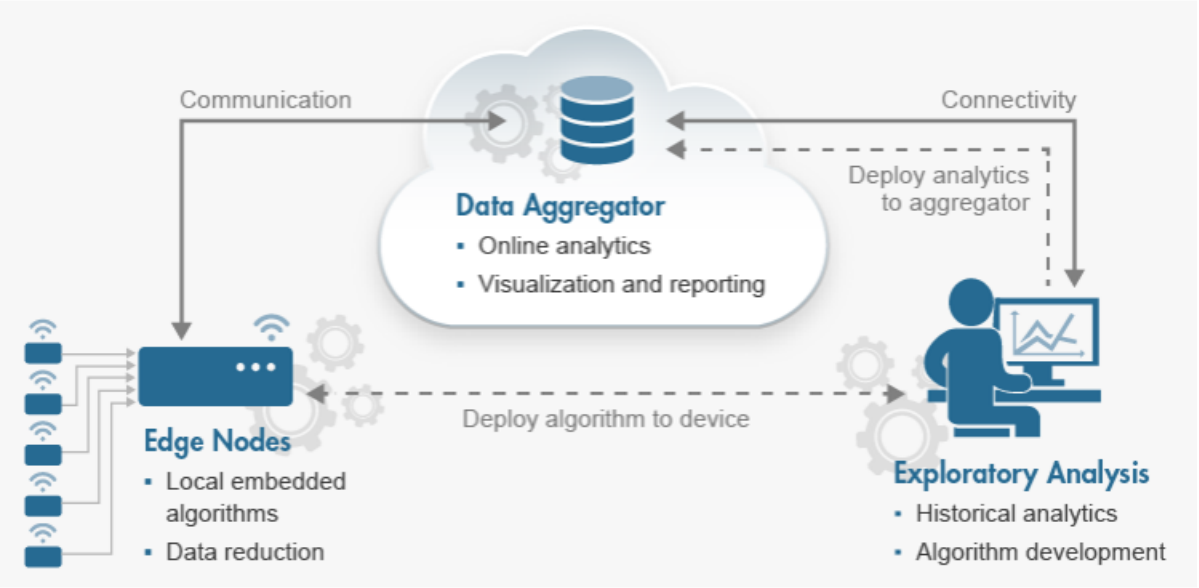
MATLAB Compiler and MATLAB Coder



The Internet of Things

What is ThingSpeak?

- Free online data aggregation platform
  - Typically used to collect data from sensors (“Things”)
  - Provides instant visualization of the data
  - Popular for people experimenting in IoT
- Can be used to act on data
  - E.g. Tweet a message when the temperature in your backyard reaches 32 degrees
- Can be used to analyze data
  - MATLAB integration allows users to run scheduled MATLAB code on data coming into ThingSpeak



Collect

Send sensor data to the cloud.

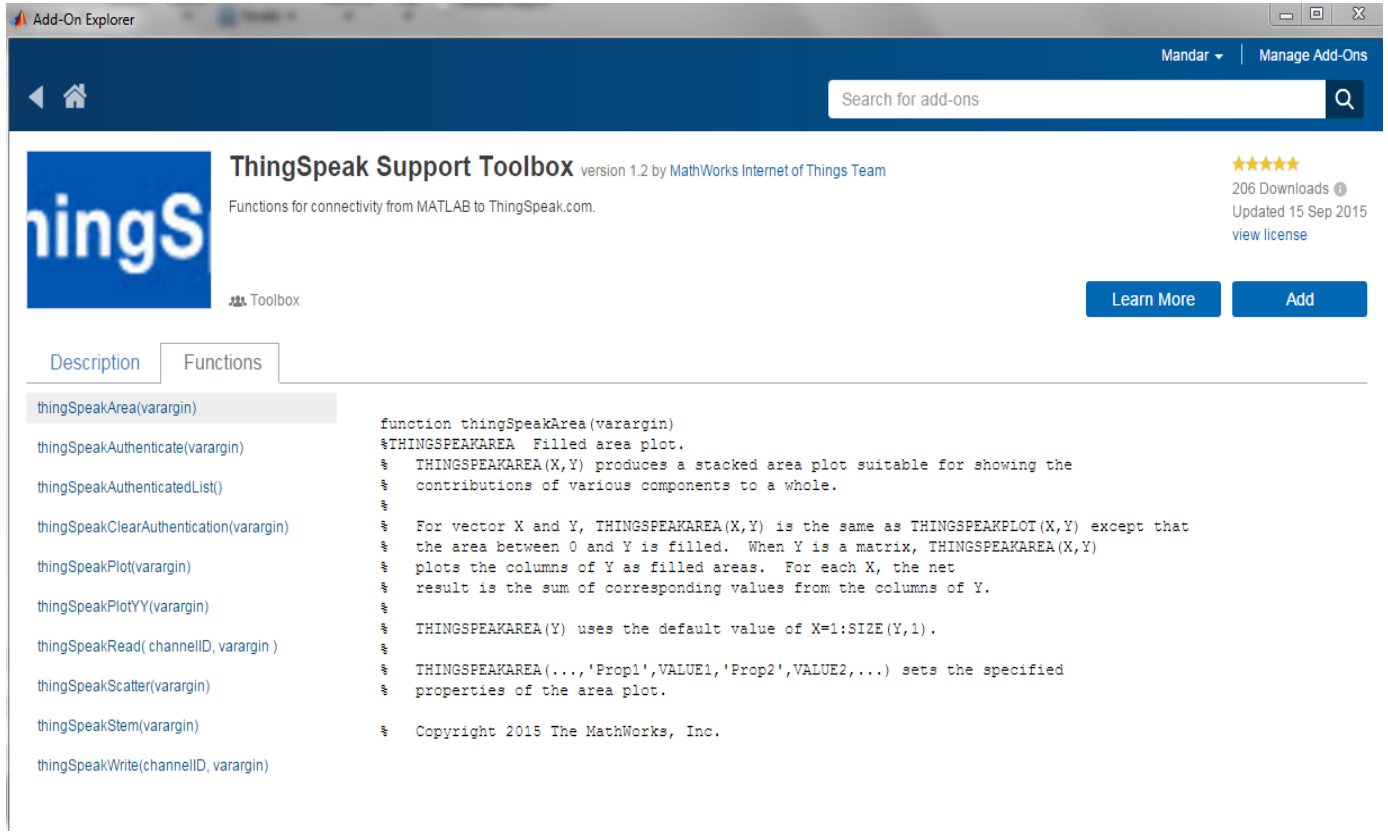
Analyze

Analyze and visualize your data.

Act

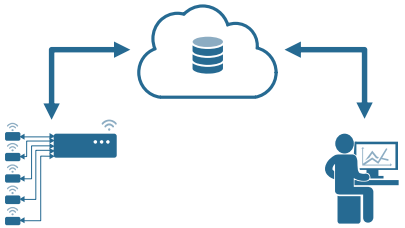
Trigger a reaction.

Support for ThingSpeak from MATLAB



ThingSpeak Weather Station

http://makerzone.mathworks.com/resources/arduino/weather-station-data-analysis/



Objectives

- Measure, explore, discover weather patterns

Solution

- Arduino station with weather sensors
- Cloud-based aggregation and analysis

https://thingspeak.com/channels/12397

In summary

1. **How can I handle diverse and/or Big Data ?**
   - *Functionalities to work with business, scientific and engineering data, connect to databases, clouds, offload computations, etc.*

2. **How to handle/create advanced algorithms, e.g., Machine Learning ?**
   - *Range of apps catered for variety of products, e.g., Machine Learning, Neural Networks, etc.*

3. **How can I integrate these algorithms?**
   - *MATLAB Compiler and Coder products help to integrate this with enterprise and business systems. Also support to low cost hardware.*

Let's speak MATLAB !

- MATLAB is Trusted
  - MATLAB Central
  - User Stories - Industry
  - User Stories – Academia
- Interested to Learn?
  - MATLAB Training
  - Get started with our free online tutorial.
  - Join our training courses for Statistics and Machine Learning.
- Speak to us, Mandar.Gujrathi@mathworks.com.au