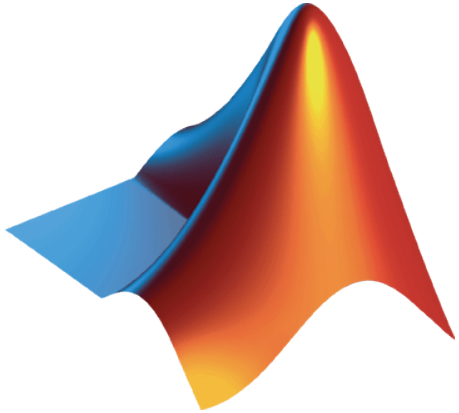
# Machine Learning for Financial Applications



Dr. Alexander Diethert, Application Engineer

Opportunities

# ~~Challenges~~ for Analysts

**How to work with your data to**

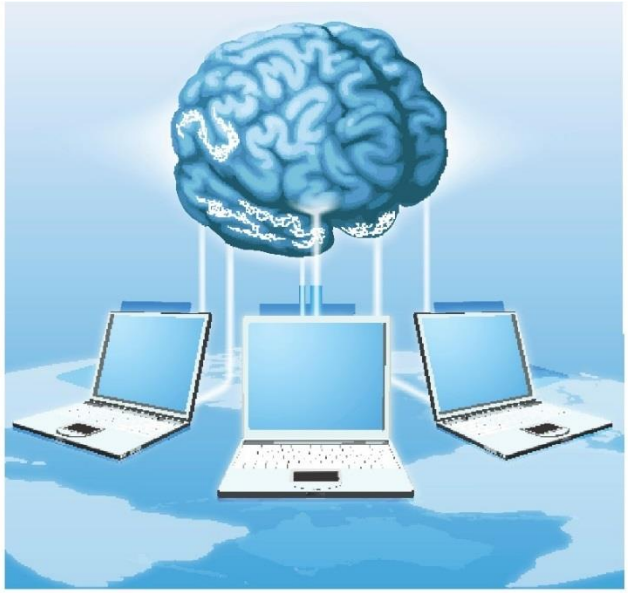
… analyze for insights?

… identify patterns?

… develop predictive models?

…deploy to collaborators?

**… innovate?**



# Customer Examples

**Machine Learning and Visualisation in the Context of a Large Enterprise**
Arjun Viswanathan, Citi

## MATLAB Used to Predict Financial Crises in Emerging Markets

"Because MATLAB is both powerful and easy to use, I felt confident that the Bank of Indonesia would be able to implement the MATLAB programs and use them as an early warning system for financial distress."

— *Dr. Paul McNelis, Georgetown University*

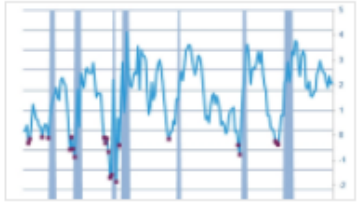
**Machine Learning and Applications in Finance**
Christian Hesse, Deutsche Bank and University College London

## Gas Natural Fenosa Predicts Energy Supply and Demand

"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*

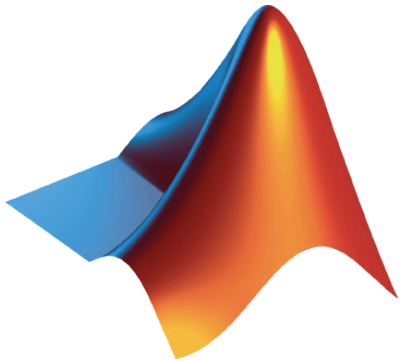
Portomouros hydroelectric dam.



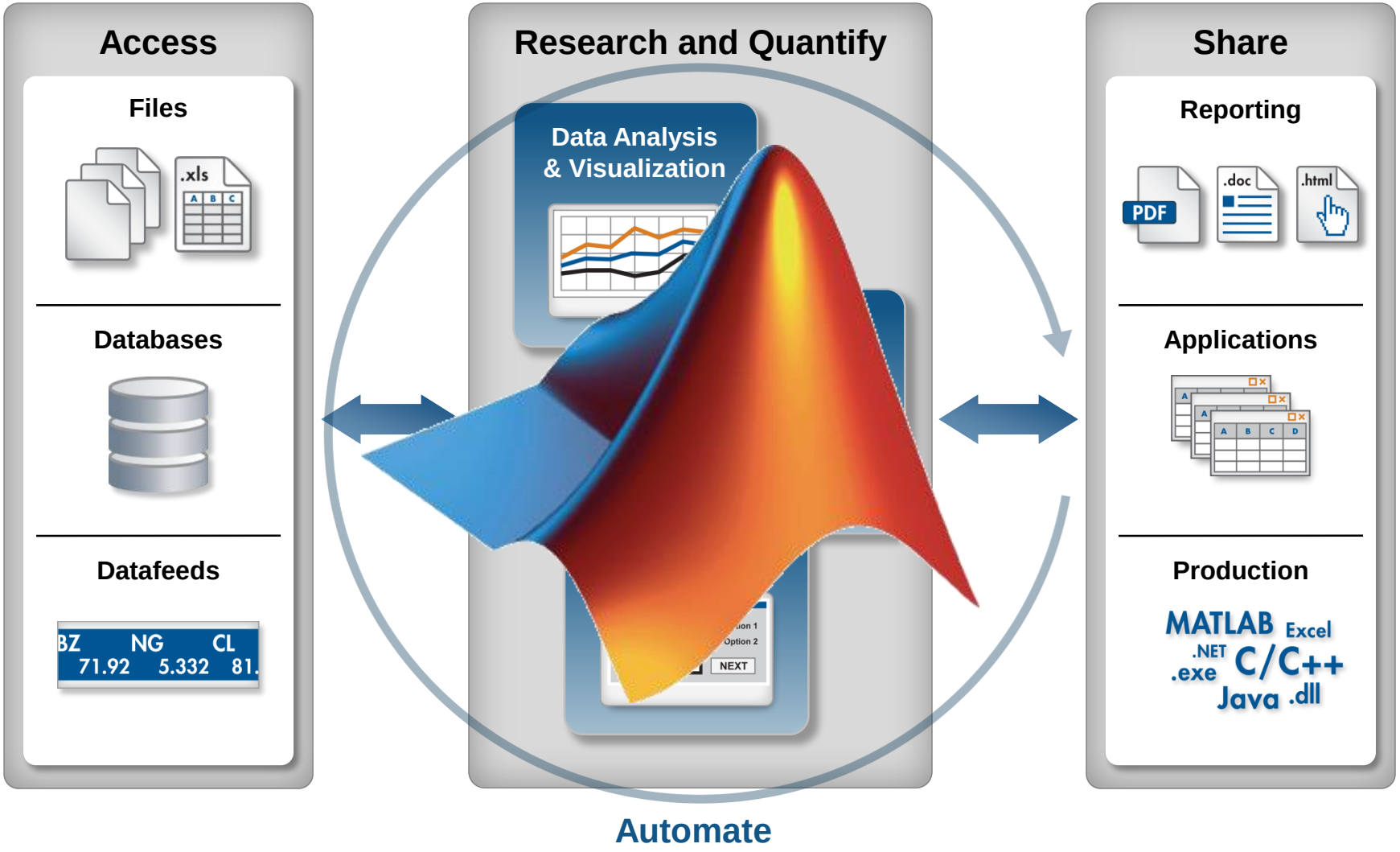
**Portfolio Allocation with Machine Learning and MATLAB Distributed Computing Server on Microsoft Azure Cloud**
Emilio Llorente-Cano and James Mann, Aberdeen Asset Management

# Agenda

- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary

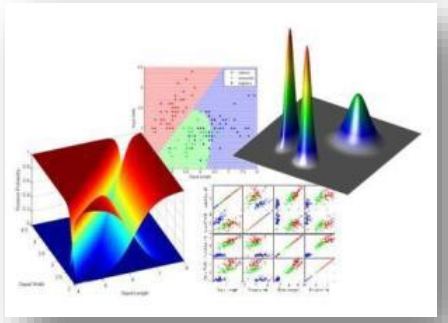


# Computational Finance Workflow

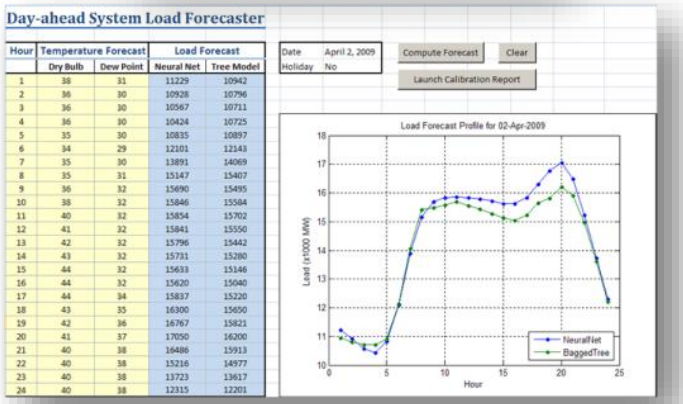


# Machine Learning

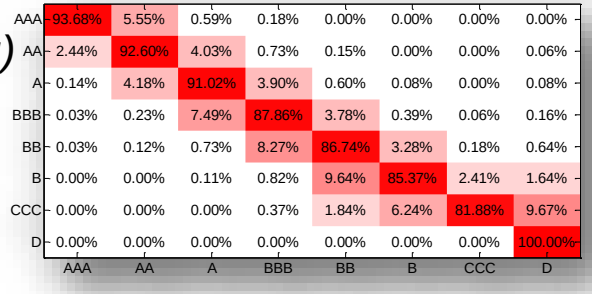
## Characteristics and Examples



- Characteristics
  - Lots of data (many variables)
  - System too complex to know the governing equation *(e.g., black-box modeling)*



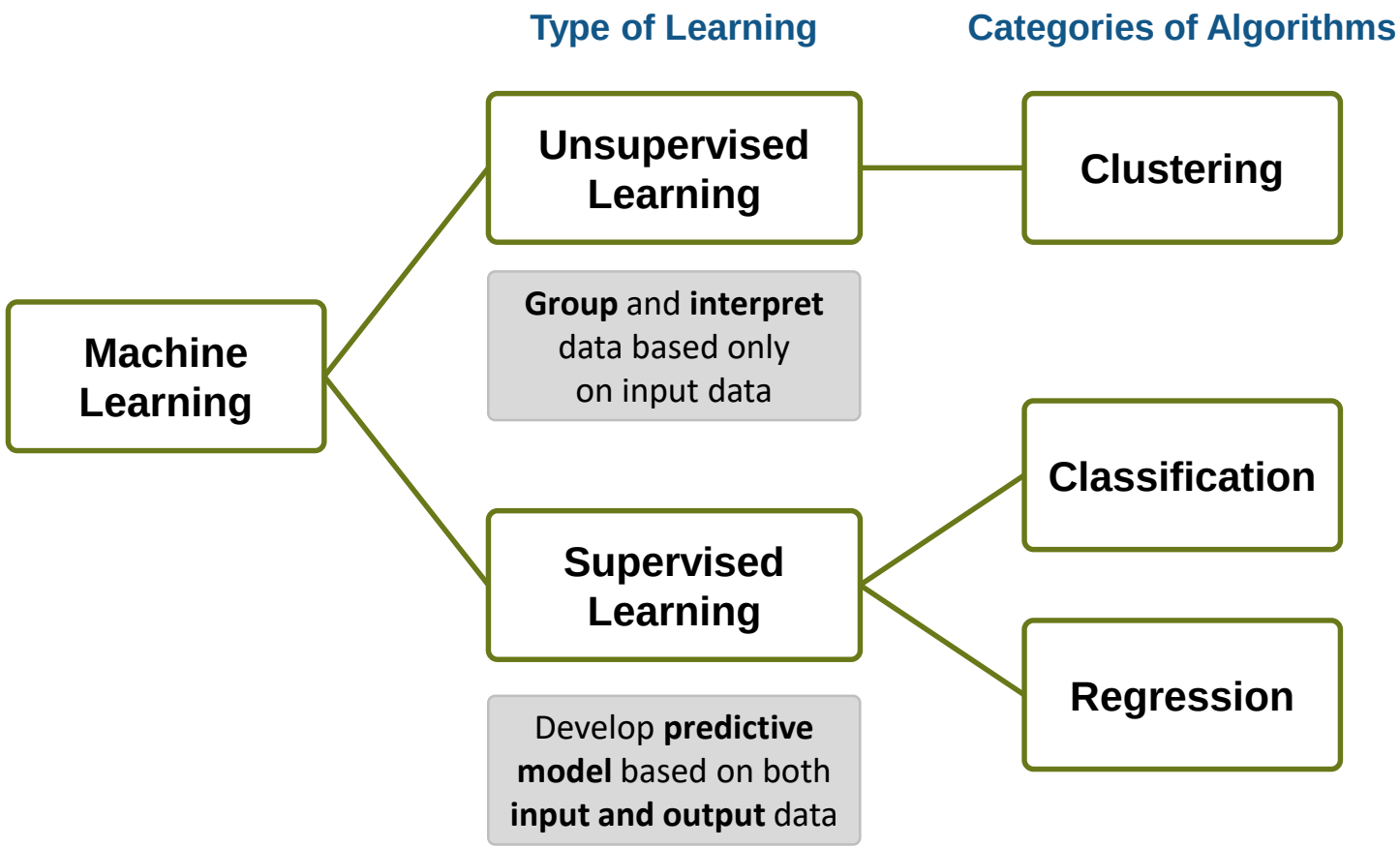

- Examples
  - Pattern recognition *(speech, images)*
  - Financial algorithms *(credit scoring, algo trading)*
  - Energy forecasting *(load, price)*
  - Biology *(tumor detection, drug discovery)*



# Challenges – Machine Learning

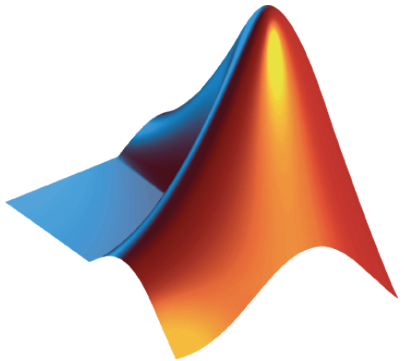
- Significant technical expertise required
- No “one size fits all” solution
- Locked into Black Box solutions
- Time required to conduct the analysis

# Overview – Machine Learning

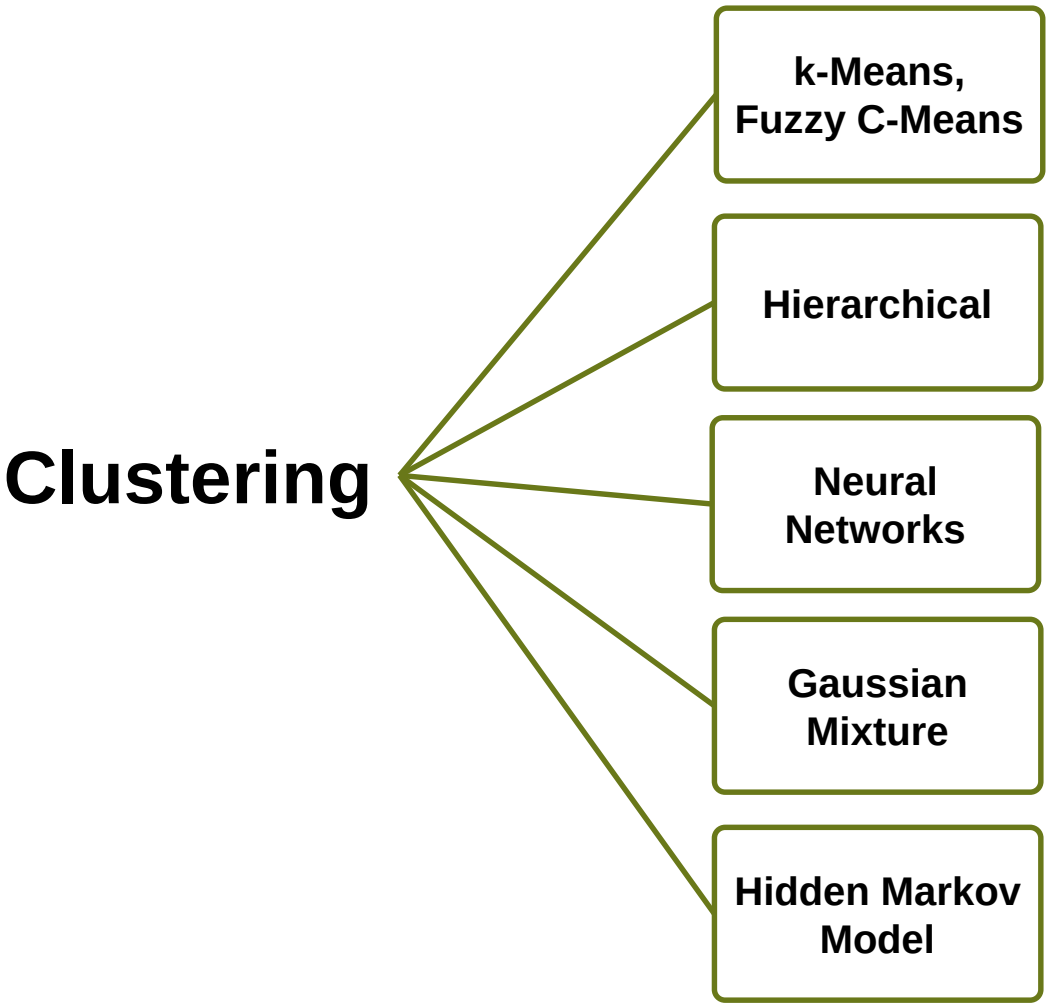


# Agenda

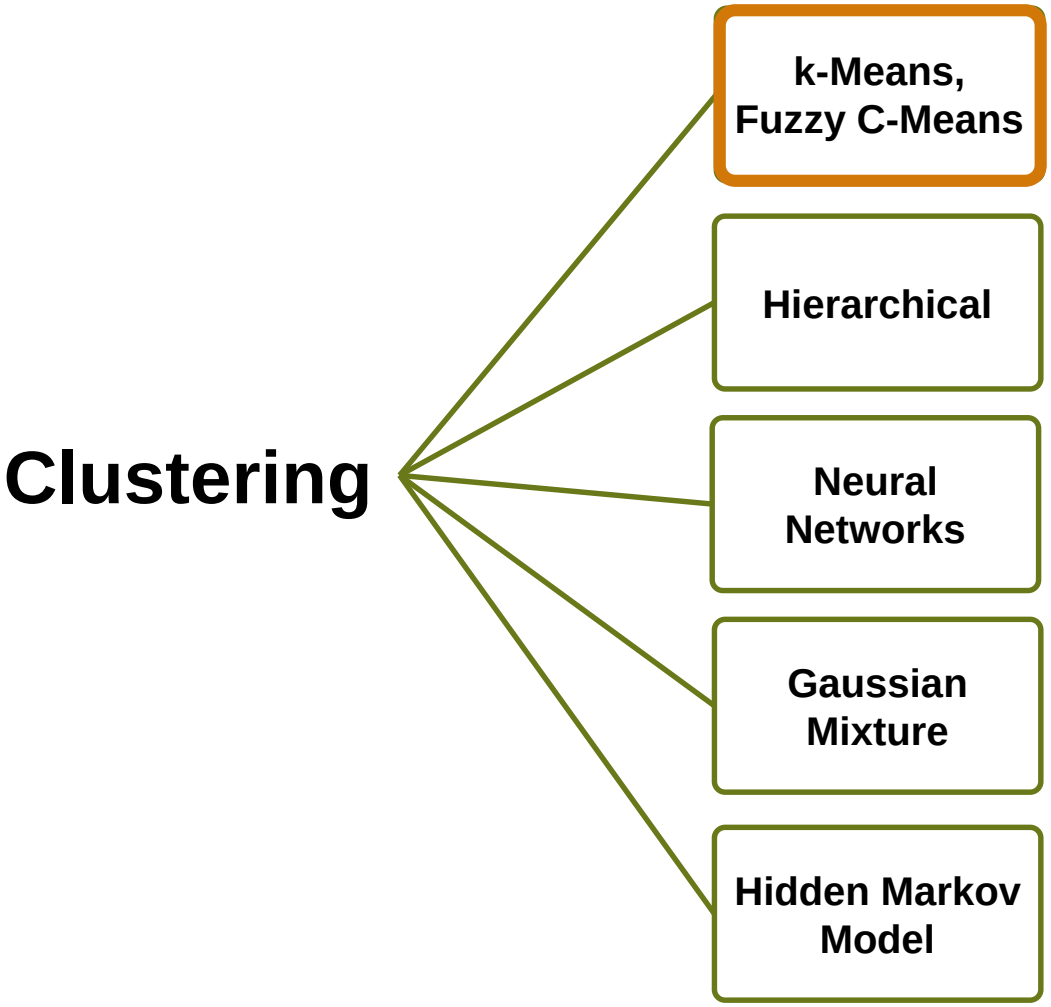
- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary



# Unsupervised Learning



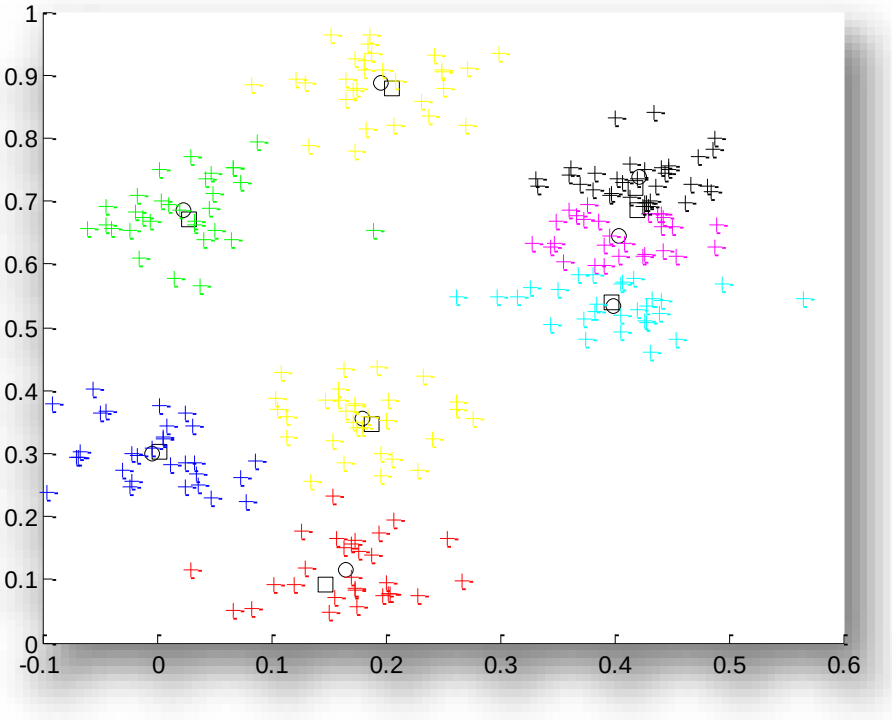
# Unsupervised Learning



# Clustering

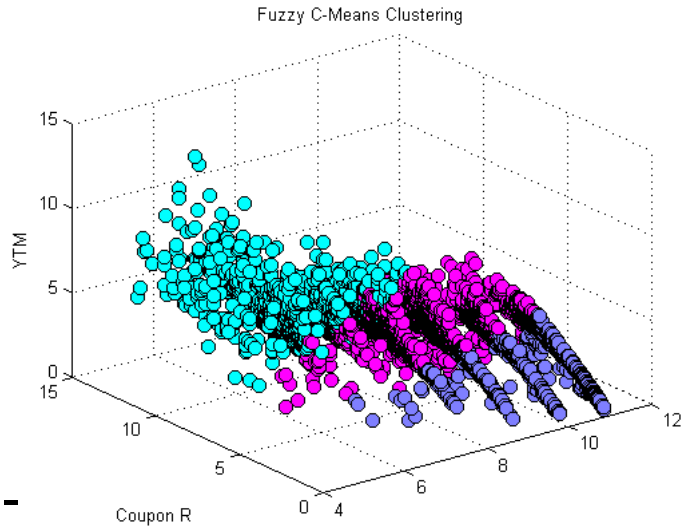
## Overview

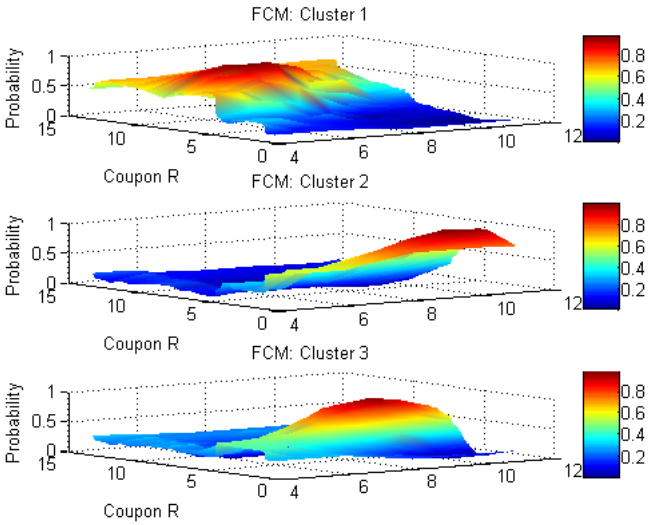
- What is clustering?
  - Segment data into groups, based on data similarity
- Why use clustering?
  - Identify outliers
  - Resulting groups may be the matter of interest
- How is clustering done?
  - Can be achieved by various algorithms
  - It is an iterative process *(involving trial and error)*



# Example – Clustering Corporate Bonds

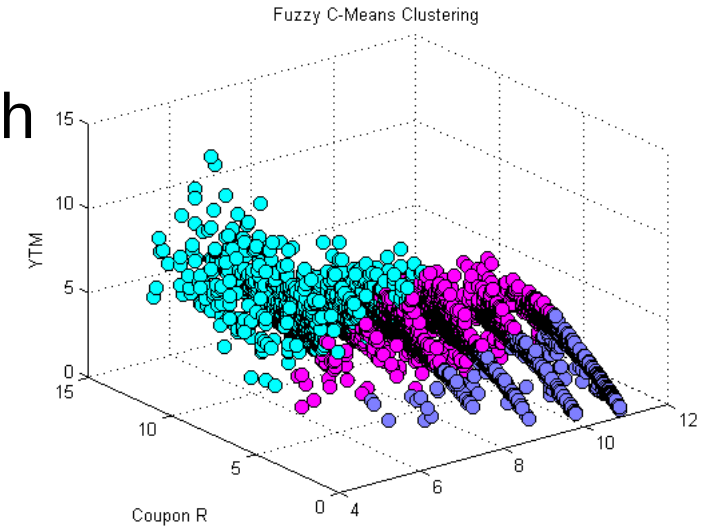
- Goal:
  - Cluster similar corporate bonds together
- Approach:
  - Cluster the bonds data using distance-based and probability-based techniques

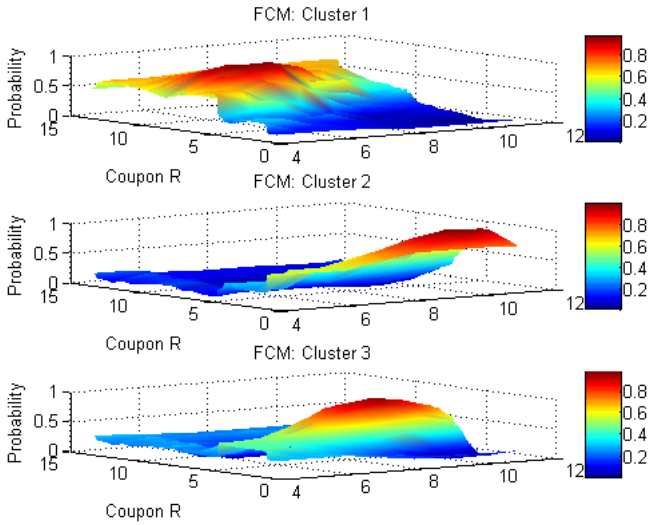


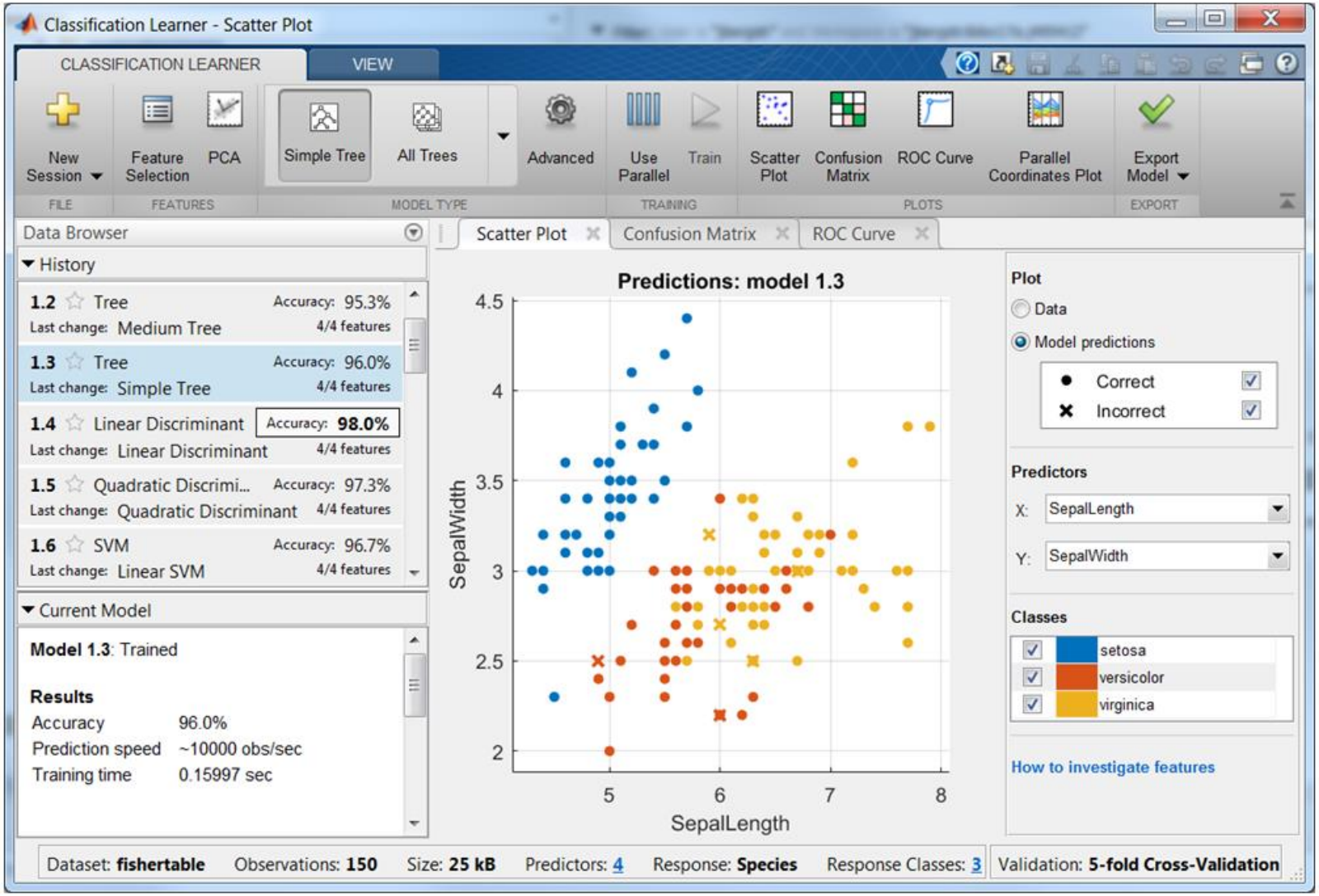
# Example – Clustering Corporate Bonds

- Numerous clustering functions with rich documentation
- Interactive visualizations to aid discovery
- Viewable source; not a black box
- Rapid exploration & development



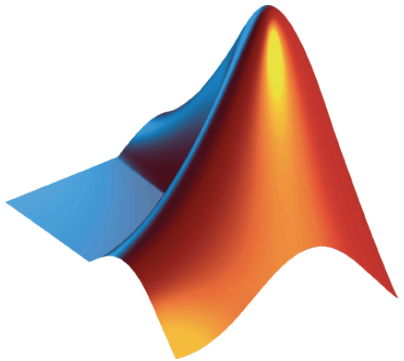



# Classification Learner App

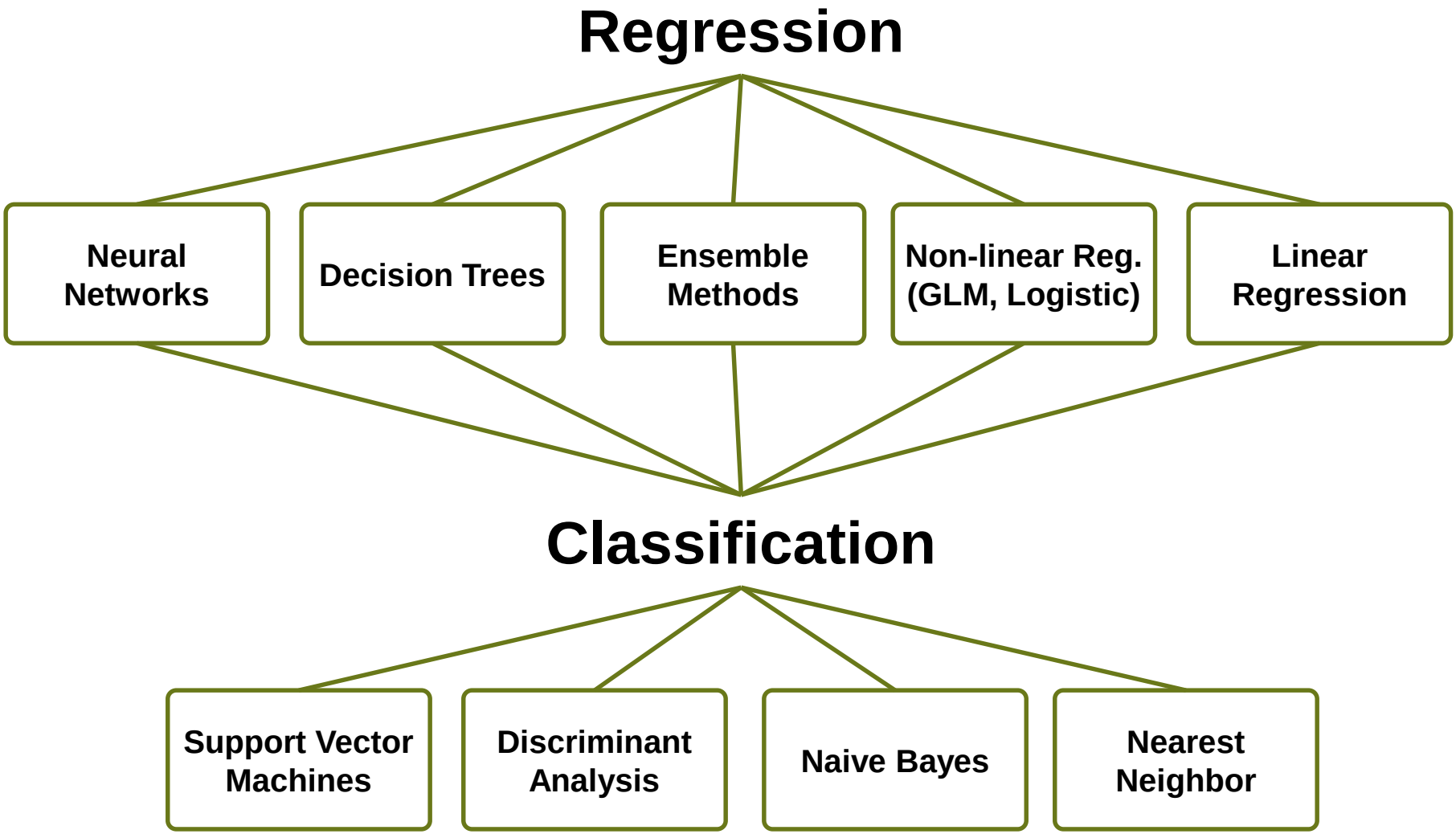


# Agenda

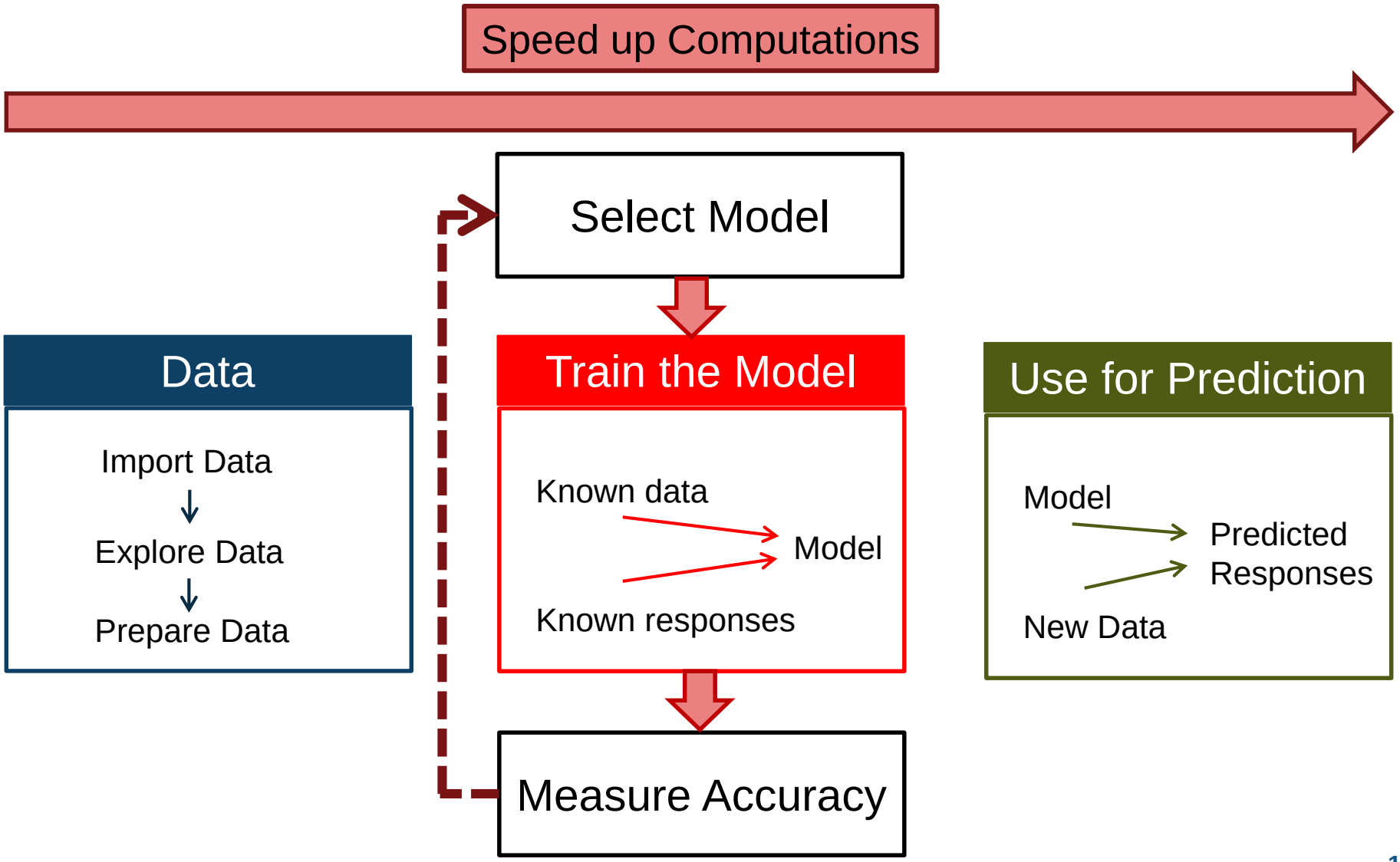
- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary



# Supervised Learning



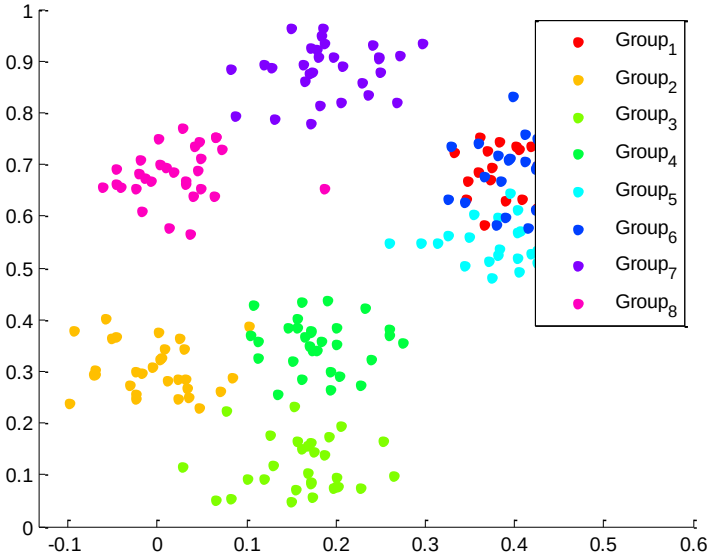
# Supervised Learning - Workflow



# Classification

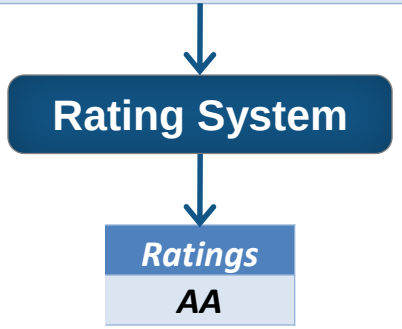
## Overview

- What is classification?
  - Predicting the best group for each point
  - “Learns” from labeled observations
  - Uses input features
- Why use classification?
  - Accurately group data never seen before
- How is classification done?
  - Can use several algorithms to build a predictive model
  - Good training data is critical



# Example: Calibrating a rating system

| ID | WC_TA | RE_TA | EBIT_TA | MVE_BVTD | S_TA | Industry |
|---|---|---|---|---|---|---|
| 60644 | 0.049 | 0.220 | 0.041 | 2.400 | 0.489 | 6 |

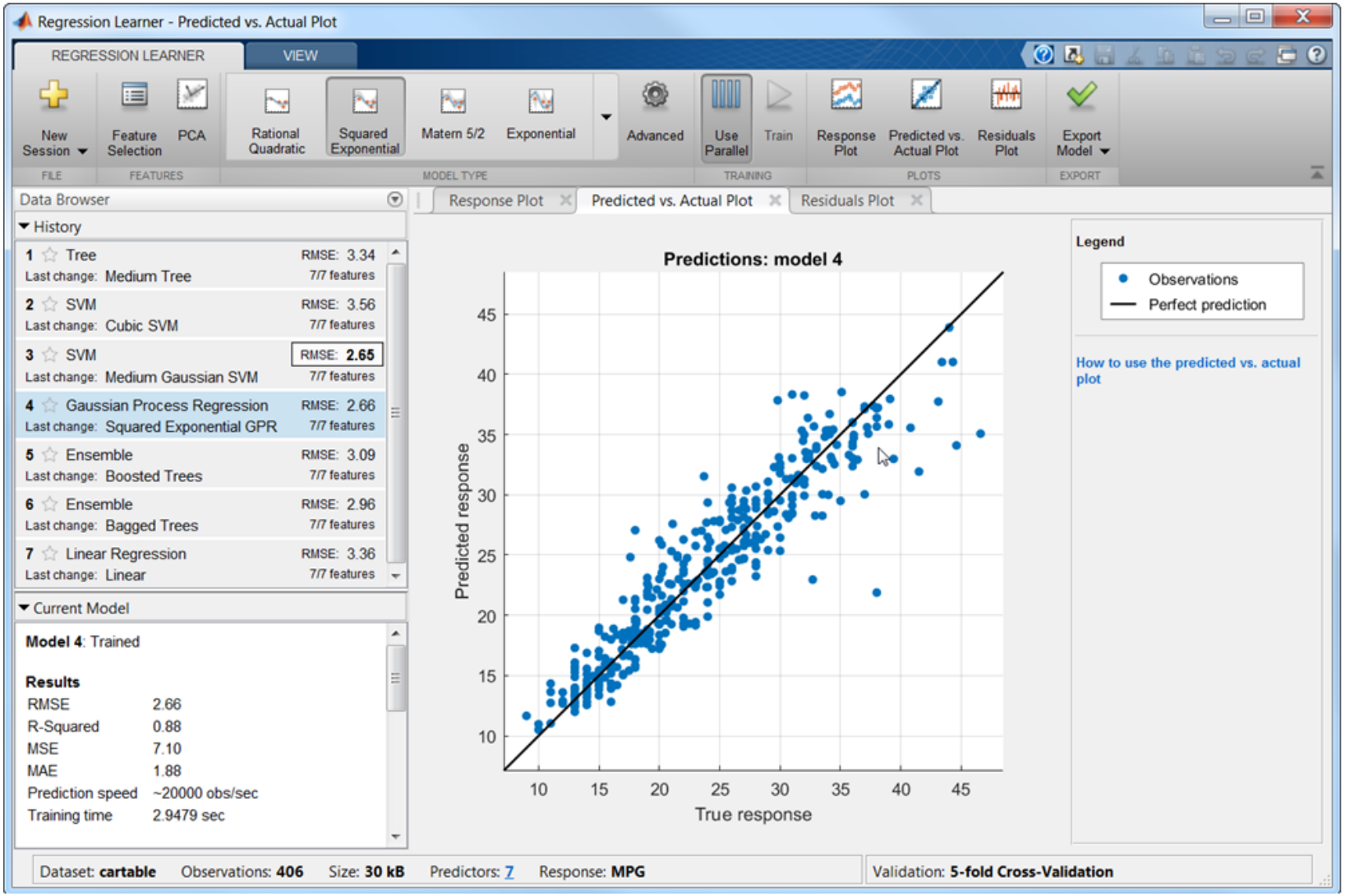



| Ratings |
|---|
| AA |

## Key tasks

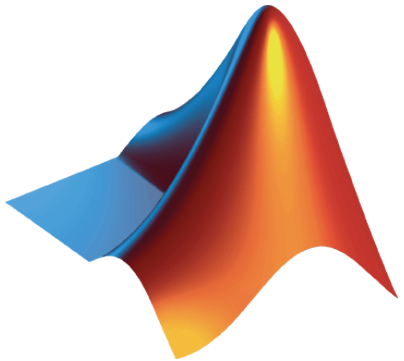
1. Perform *ad hoc* data analysis (linear regression)
2. Improve existing analysis (logistic regression)
3. Create a robust classifier: TreeBagger

# Regression Learner App



# Agenda

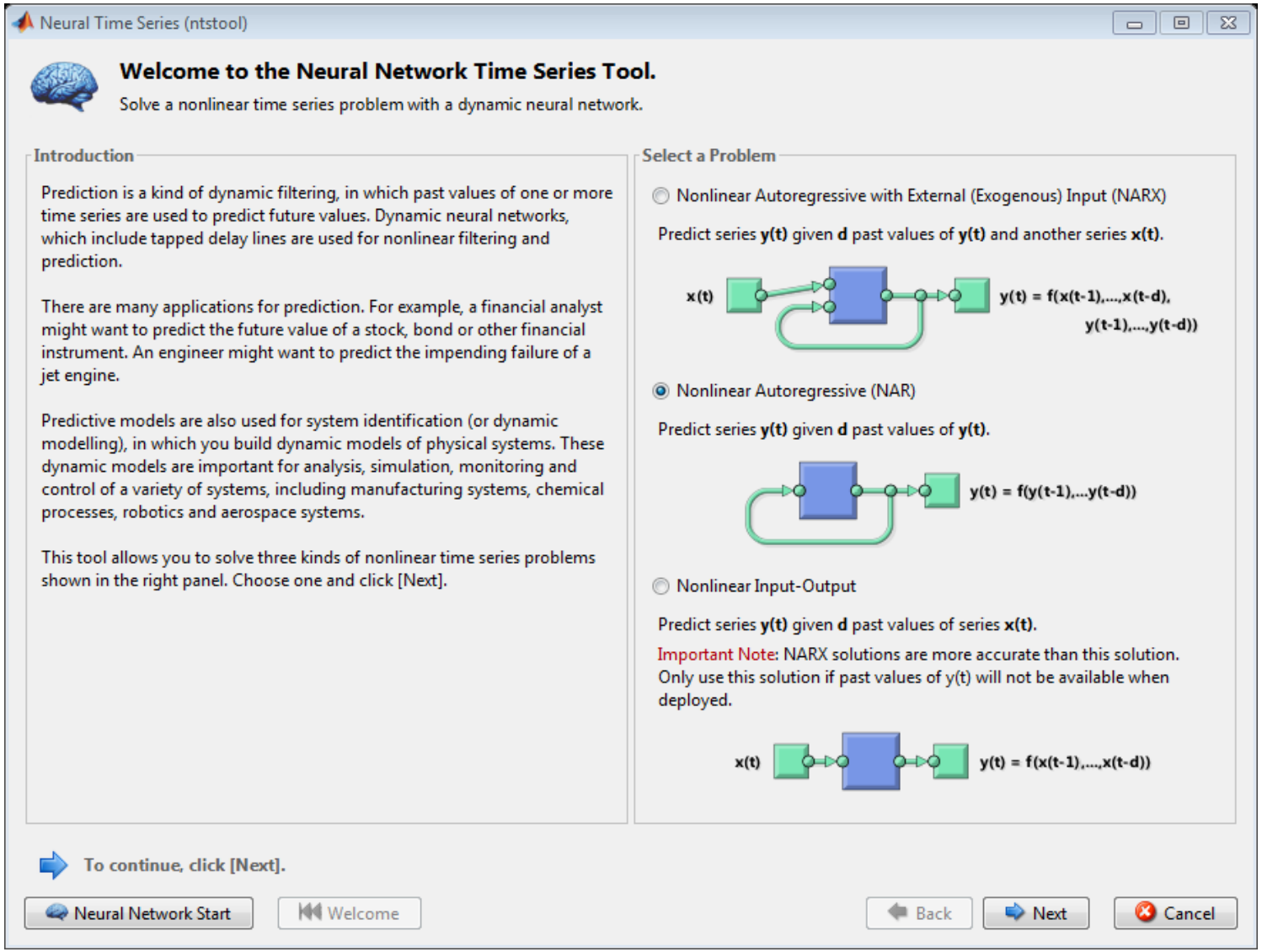
- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary



# Neural Network Use Cases

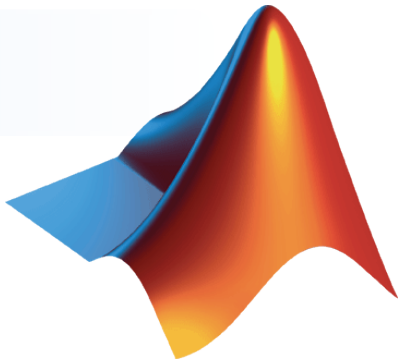
- Function Approximation and Nonlinear Regression
- Pattern Recognition and Classification
- Clustering
  - Self-Organizing Maps
  - Competitive Layers
- Time Series and Dynamic Systems
  - Modeling and Prediction with NARX and Time-Delay Networks
  - Creating Simulink Models

# Neural Network Apps

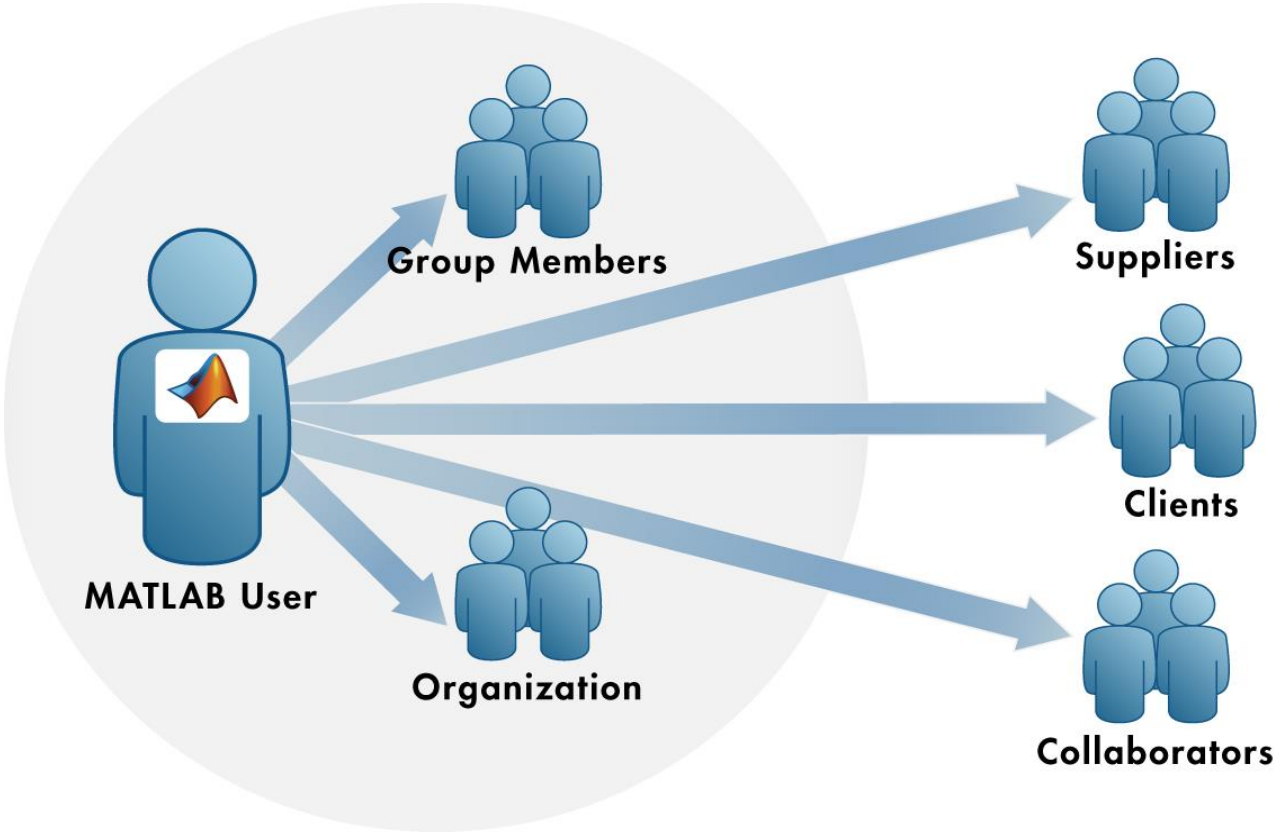


# Agenda

- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary

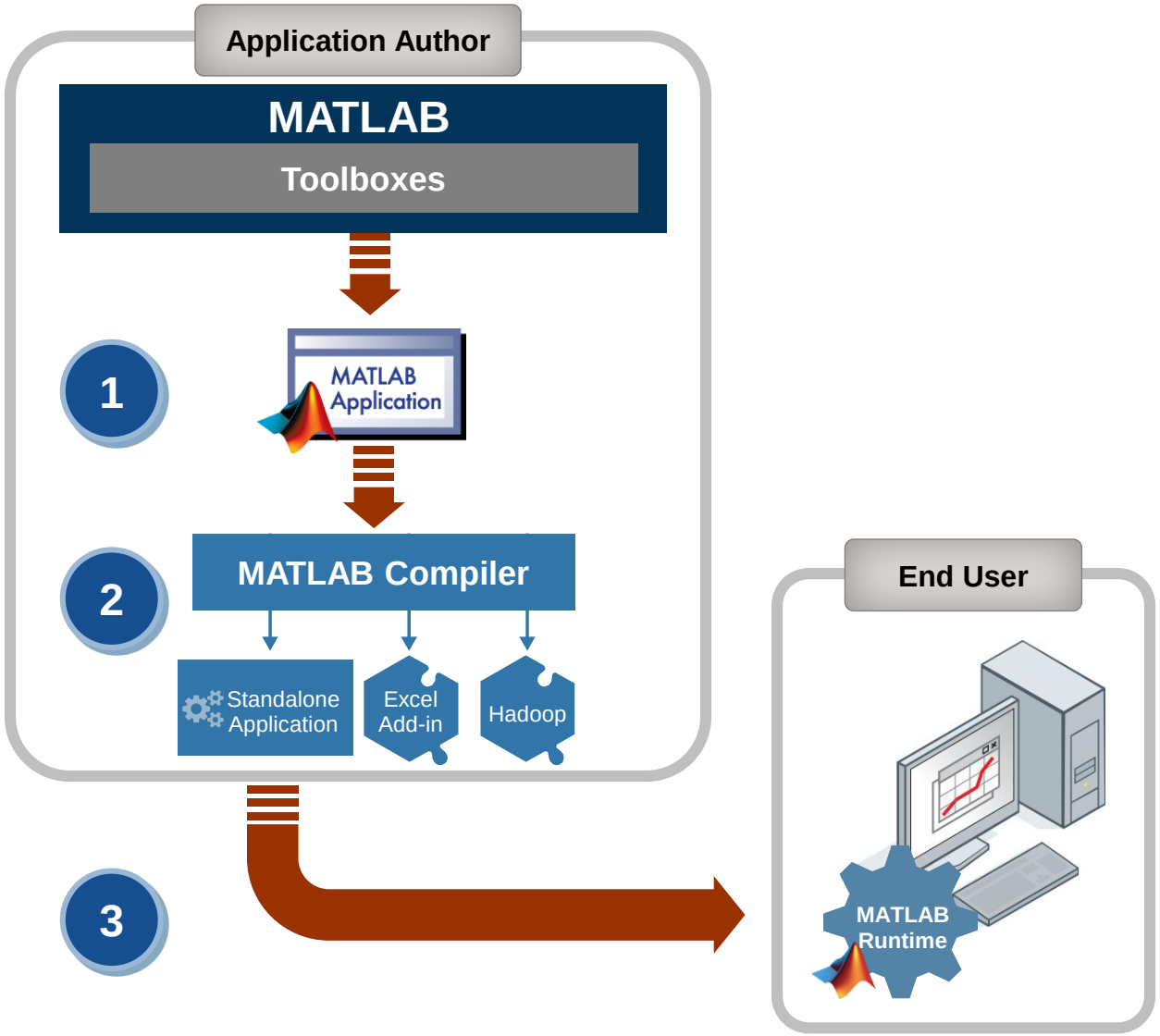


# Share Programs Outside of MATLAB

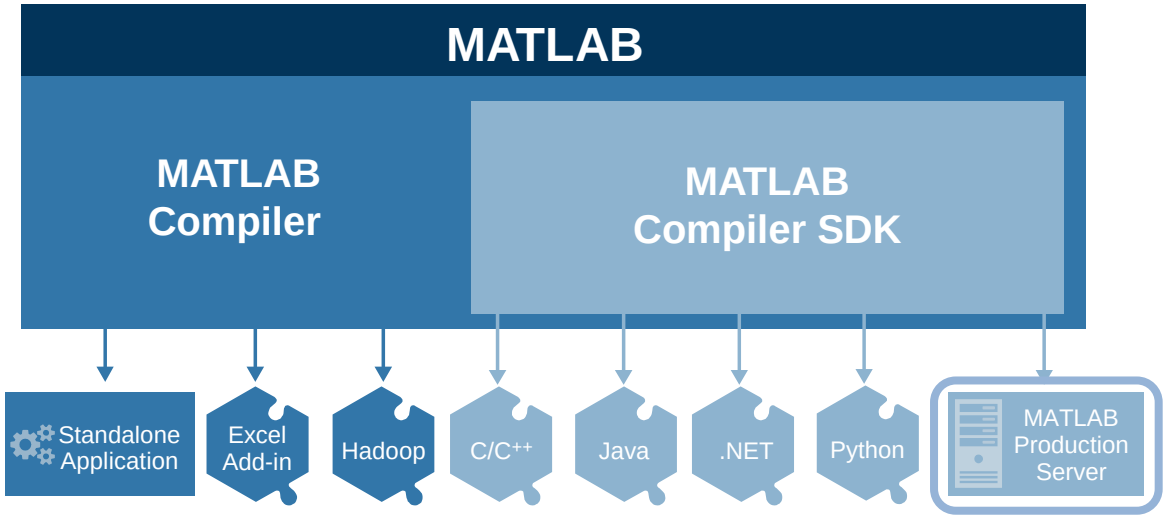



Deploy your MATLAB code to people who don't need MATLAB

# Sharing Standalone Applications



# Which Product will Fit Your Needs?

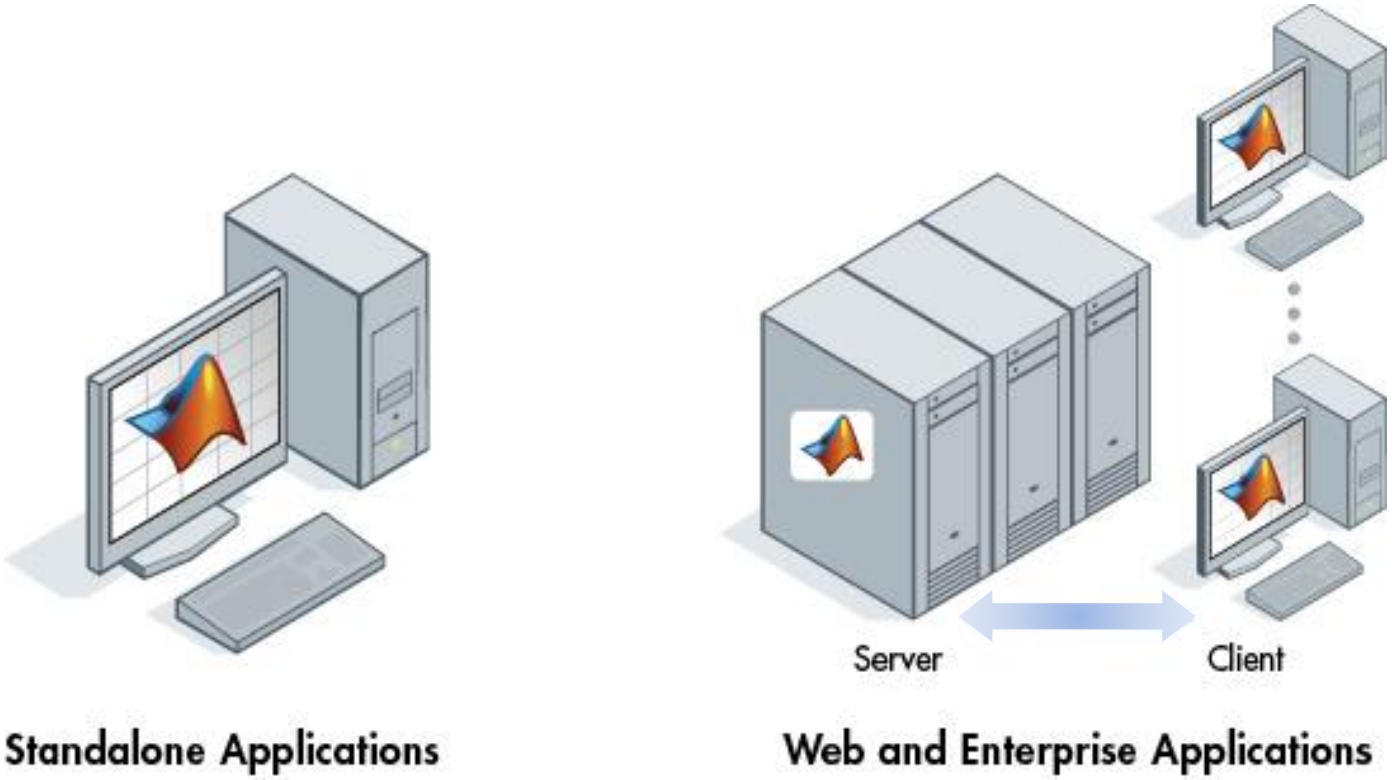



***MATLAB Compiler*** for sharing MATLAB programs without integration programming

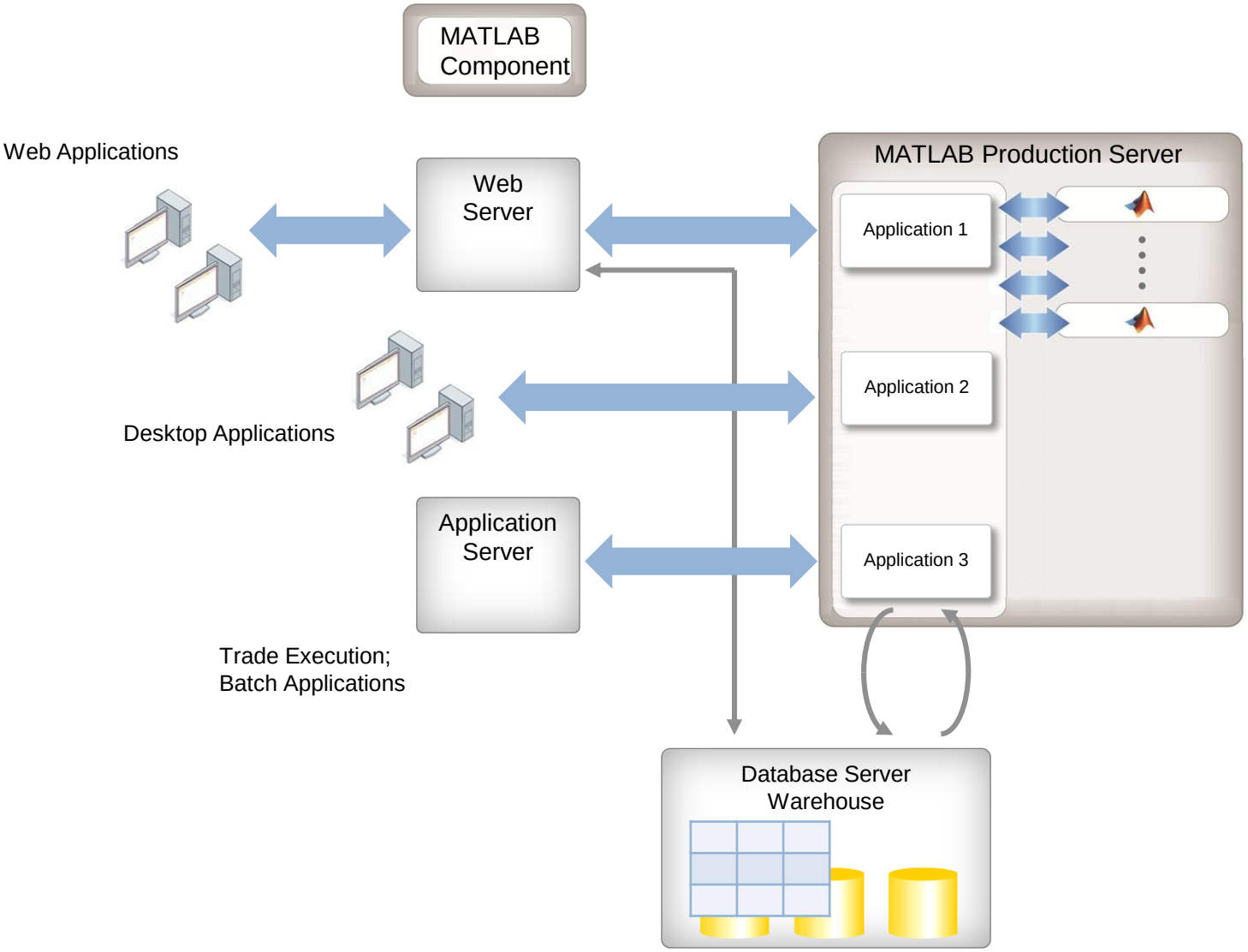
***MATLAB Compiler SDK*** provides implementation and platform flexibility for software developers

***MATLAB Production Server*** provides the most efficient development path for secure and scalable web and enterprise applications

# The Range of Application Platforms

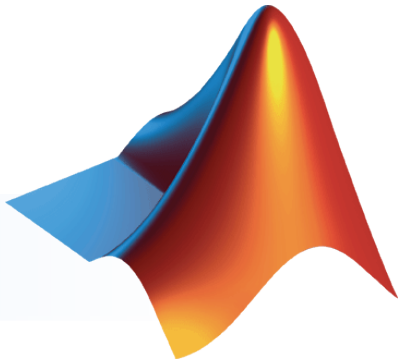


# MATLAB Production Server



# Agenda

- Introduction
- Unsupervised Learning: Clustering Bond Data
- Supervised Learning: Create Rating System
- Neural Network Based Time Series Forecasting
- Outlook: Distributing MATLAB Applications
- Summary



# MATLAB for Machine Learning

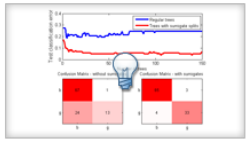
| Challenges | MATLAB Solution |
|---|---|
| Time (loss of productivity) | **Rapid analysis and application development**<br>High productivity from data preparation, interactive exploration, visualizations. |
| Extract value from data | **Machine learning, Video, Image, and Financial**<br>Depth and breadth of algorithms in classification, clustering, and regression |
| Computation speed | **Fast training and computation**<br>Parallel computation, Optimized libraries |
| Time to deploy & integrate | **Ease of deployment and leveraging enterprise**<br>Push-button deployment into production |
| Technology risk | **High-quality libraries and support**<br>Industry-standard algorithms in use in production<br>Access to support, training and advisory services when needed |

# Learn More: Machine Learning with MATLAB

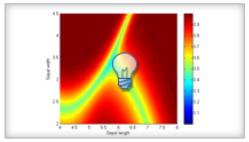
mathworks.com/machine-learning

## Classification Examples

Basket Selection Using Stepwise Regression



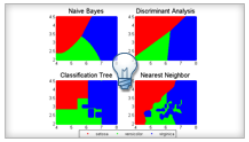
Classification in the Presence of Missing Data



Classification Probability

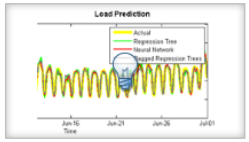
Digit Classification Using HOG Features

Handwriting Recognition Using Bagged Classification Trees

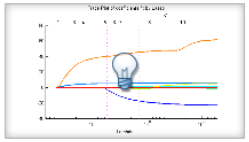


Visualize Decision Surfaces for Different Classifiers

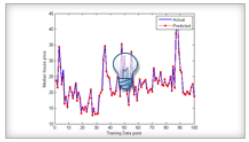
## Regression Examples



Electricity Load Forecasting

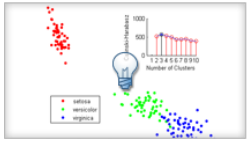


Lasso Regularization



Regression with Boosted Decision Trees

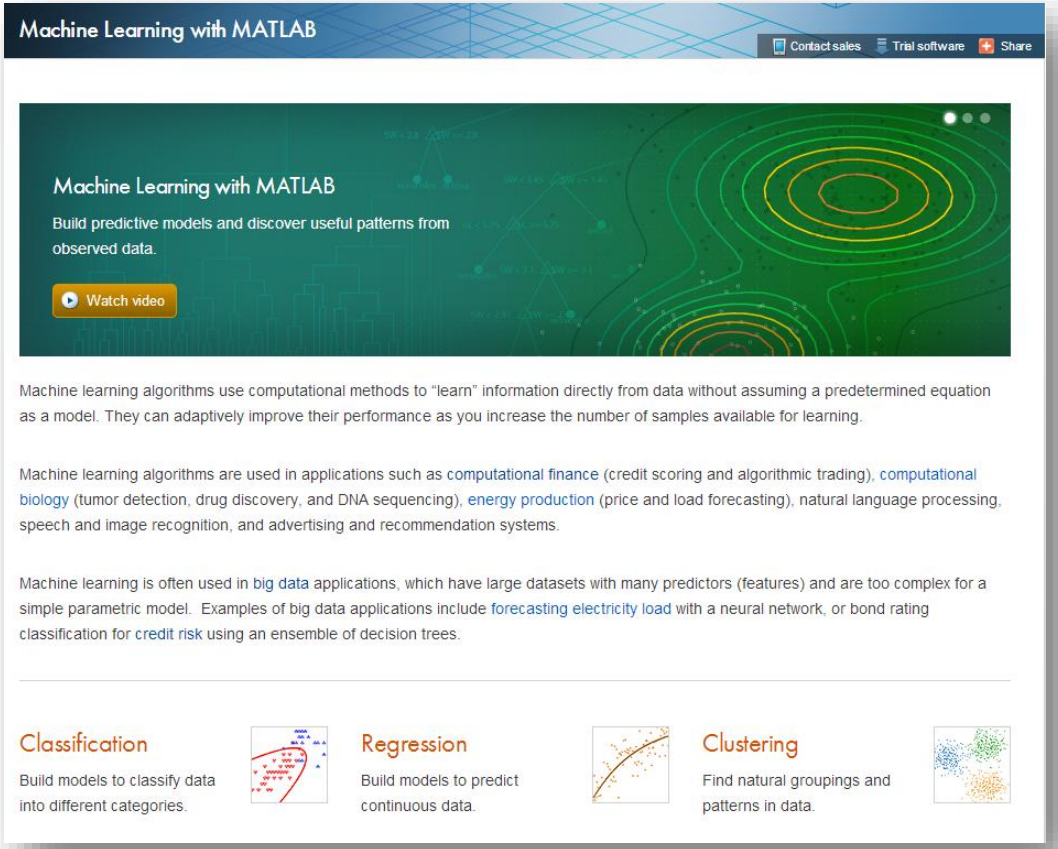
## Clustering Examples



Cluster Evaluation

Cluster Genes Using K-Means and Self-Organizing Maps

Color-Based Segmentation Using K-Means Clustering



# Thank you!