Tutorial T4: Practical Data Mining

PHM Conference, Minneapolis, MN
9/23/2012
1:30-3:00 PM

Ravi Patankar

Principal Engineer
Honeywell Aerospace
Phoenix, AZ
ravindra.patankar@honeywell.com

Objective

Honeywell

- **Getting one started with data mining**

using RapidMiner

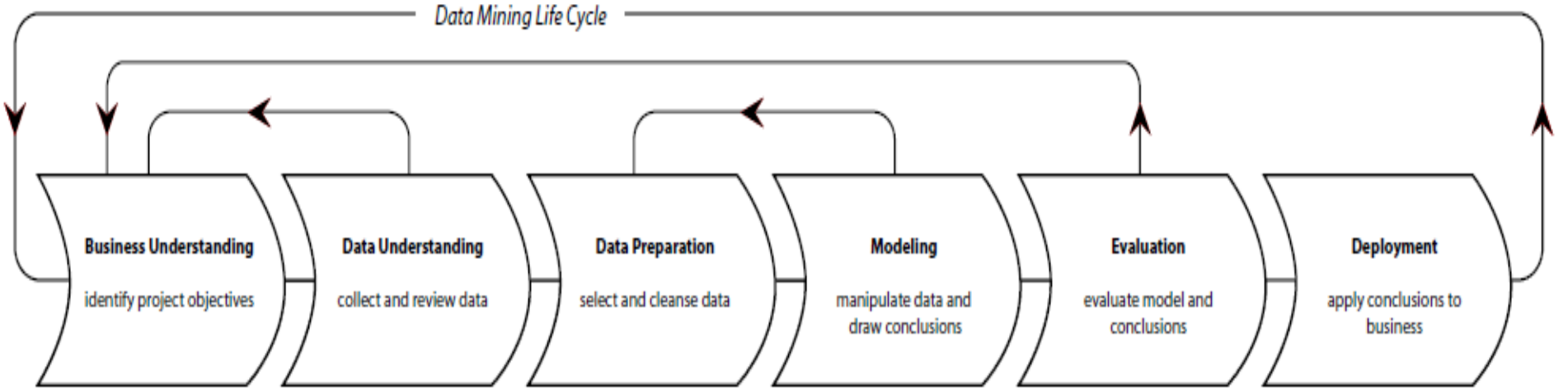
Download RapidMiner from Rapid-i.com and install

RapidMiner

Honeywell

- **Leading open source data mining tool**
- **100% java**
- **Uses CRISP-DM**

From crisp-dm.org



Agenda

Honeywell

- **RapidMiner Overview**
  - **Operators**
  - **Repositories**
    - Data
    - Process
      - (remote) Execution
      - Saving
  - **Data visualization**
  - **Documentation?**
- **ETL**
  - **Loading from data files and databases**
  - **Data preparation**
  - **Transformations**
  - **Missing values**
  - **Filtering**
  - **Outliers**
  - **Attribute reduction**
  - **Attribute selection**

Agenda (continued)

Honeywell

- **Supervised learning**
  - **Modeling methods**
  - **Applying models**
  - **Comparing models**
  - **Performance metrics and evaluation**
  - **Validation**
  - **Automatic supervised learning (PaREn)**
- **Unsupervised Learning**
  - **Association rules**
  - **Frequent item set mining**
  - **FP-Growth**
- **Text Mining Introduction**
- **Delivering Analytics via the Web**
  - **RapidAnalytics**
    - Collaboration
    - Webservices
    - Reports

Data Visualization and ETL

Honeywell

- Titanic data set from R
- Introduction to process construction
  - Decision Tree
- ETL operators

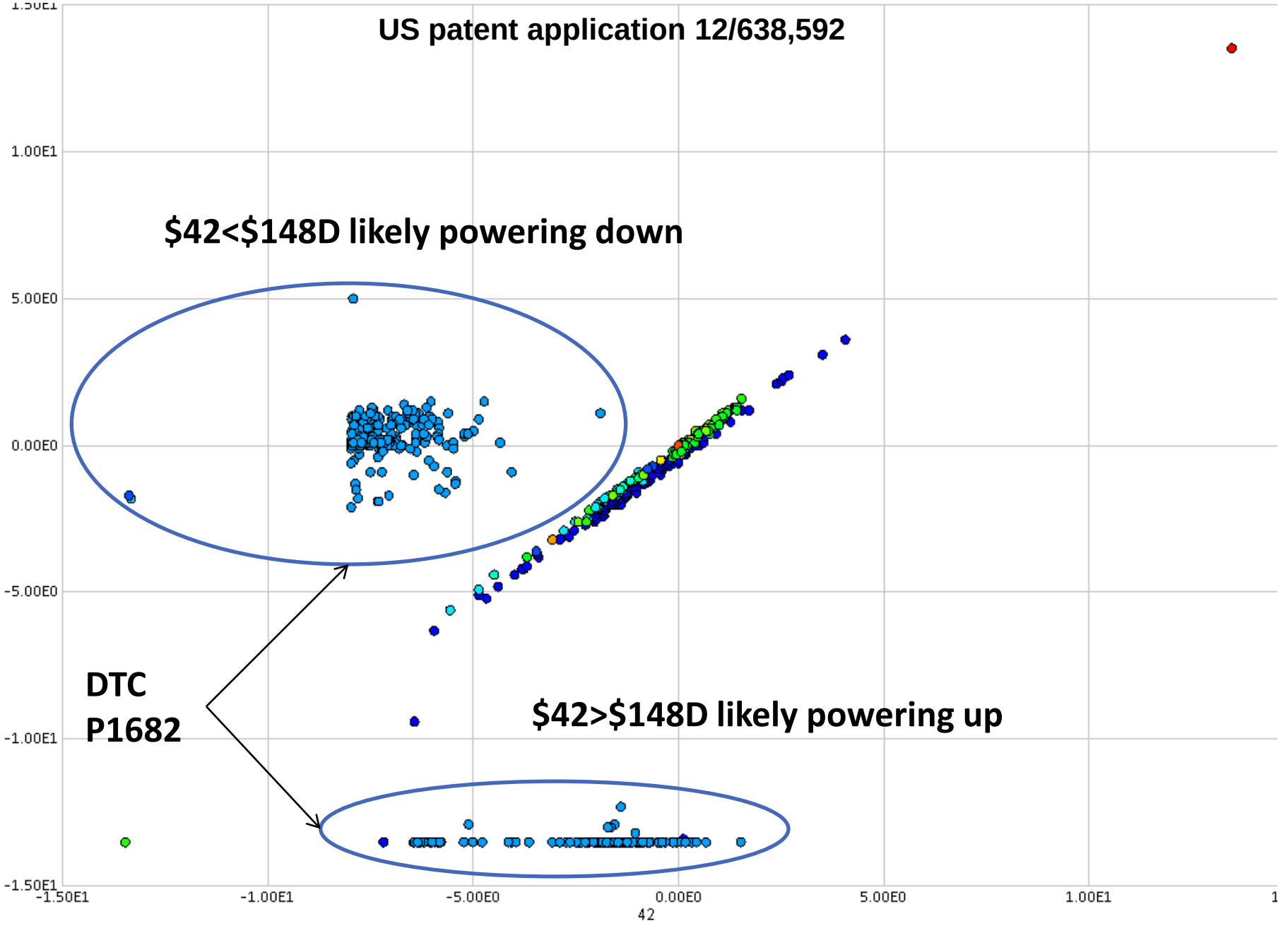
Examples

Honeywell

- Incorrect settings of automotive DTCs
- Engine X : Fuel System Failure
- Engine Y : Turbine blade and Fuel control

Control Module Vs Powertrain Voltages

Honeywell



Fuel System Related Failure

Honeywell

- **Analyze normalized engine X data**

Machine Learning

Honeywell

- **Supervised**
  - **A method for creating a function from training data**
  - **Data consist of pairs of input objects and outputs**
  - **Ex: Data mining for reasons for known fault that has occurred**
    - Algorithms: Trees - Decision Tree, Bayes, Neural Networks
- **Unsupervised**
  - **A method where a model is fit to observations.**
  - **No a priori output is known.**
  - **Ex: Data mining for patterns (faults unknown)**
    - Algorithms: Apriori, Predictive Apriori, Tertius, FPGrowth

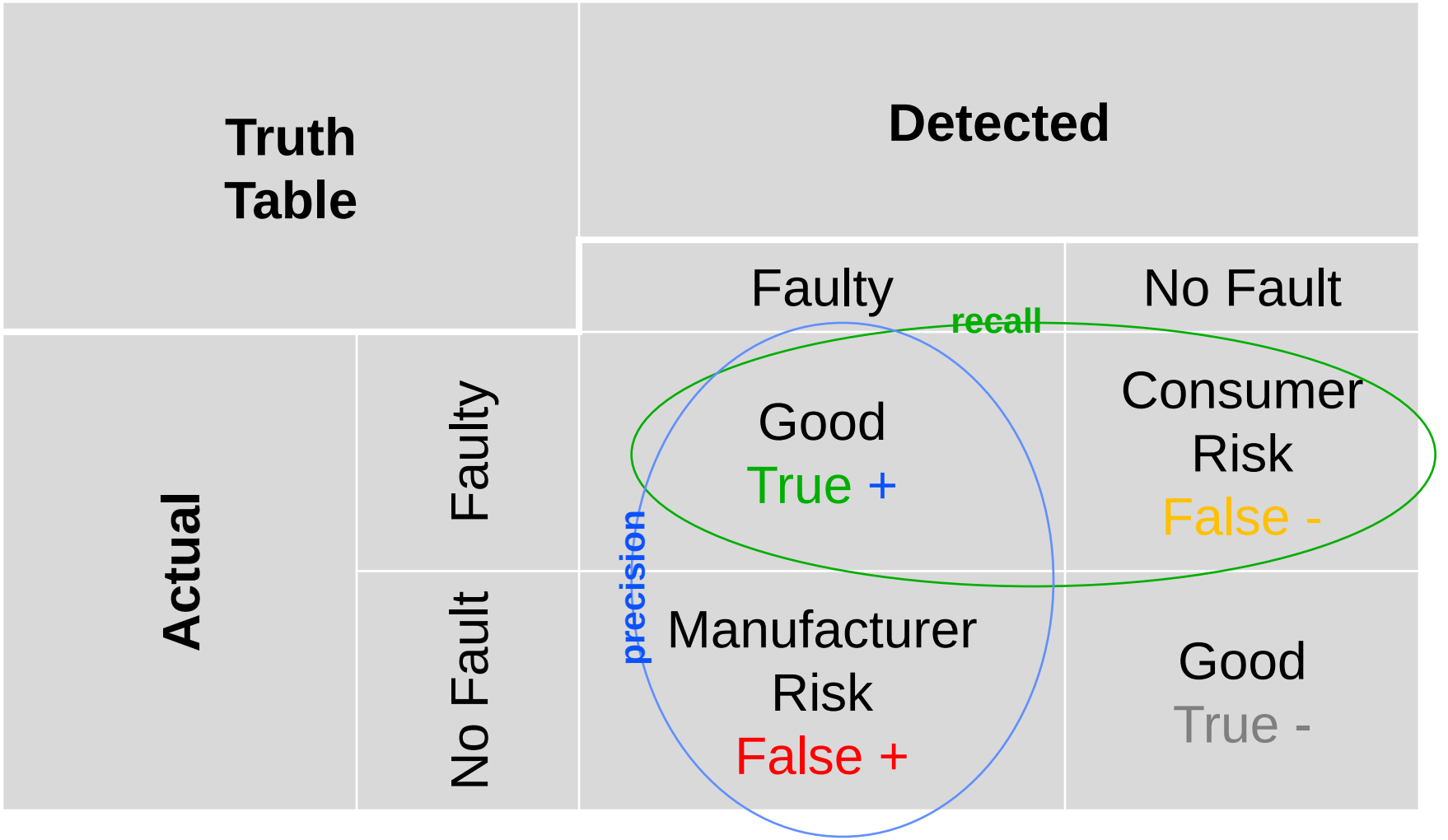
Using Supervised Learning

Honeywell

- Normalized engine Y data
- Detect Failure Onset
  - Turbine blade
  - Fuel control
- Using different Techniques
- Comparing Performance of Techniques (Methods)
- Validation Methods
- Automation/Optimization
- Using PaREn Plugin

Fault Detection Performance

Honeywell



Optimization to Tune CA-CBM

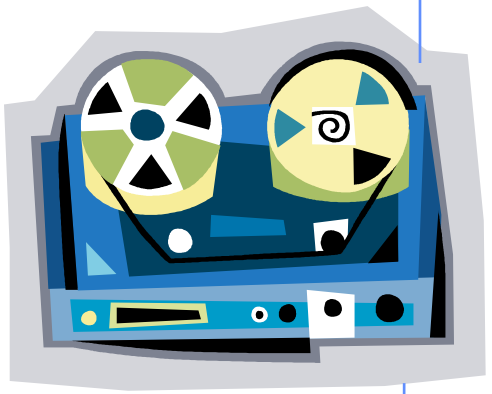
Honeywell

tunable parameters

CA-CBM input data streams

CA-CBM conclusions

Historic Data Player



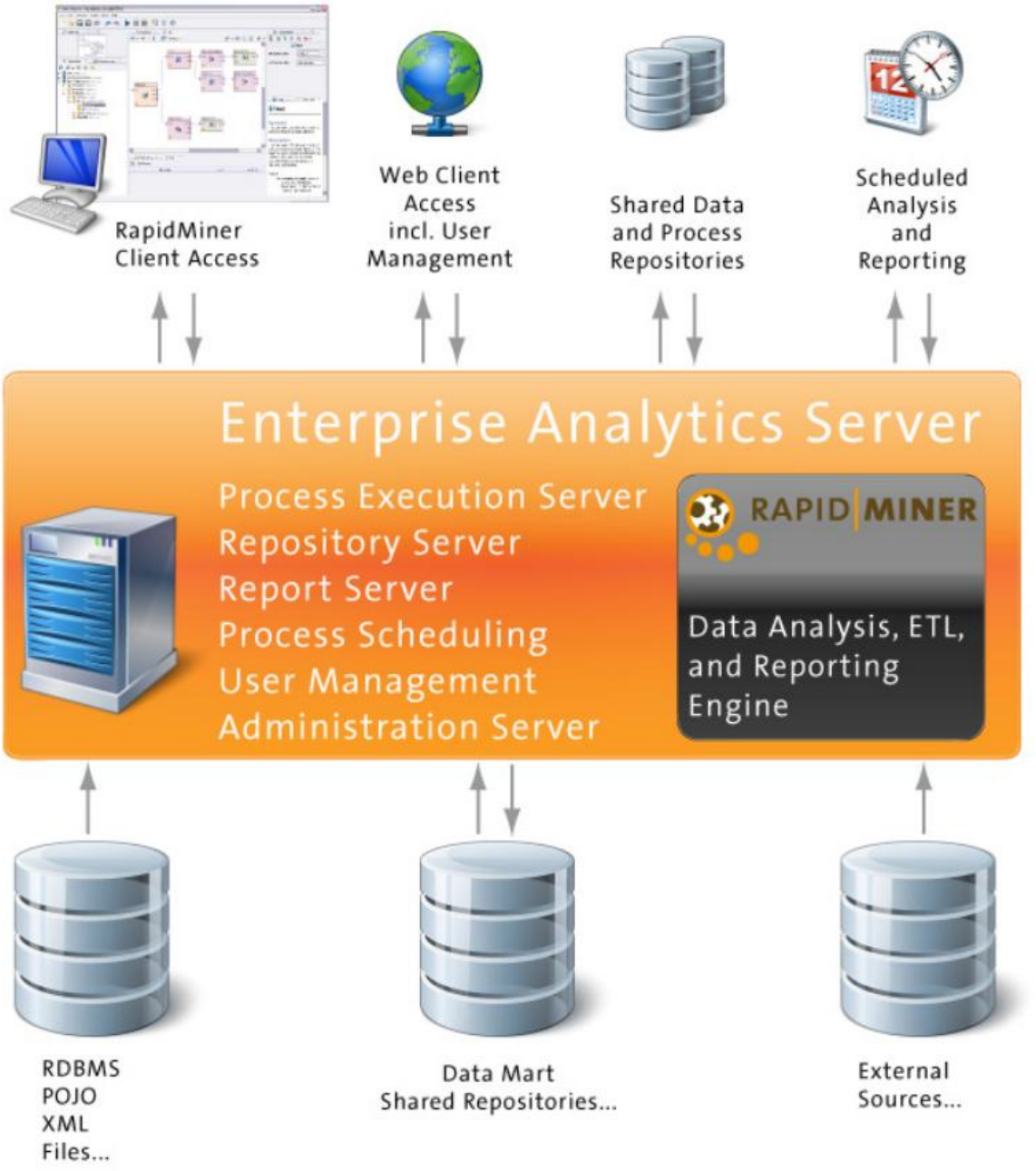
CA-CBM *performance* **Evaluator**

actual conclusions for the data streams

Optimization automatically tunes *parameters* **to obtain the maximum** *performance*

RapidAnalytics

Honeywell



Delivering Analytics via the Web

Honeywell

- **Deployment of Analyses with RapidAnalytics**
  - Sharing data, models, and processes
  - Managing processes and services
  - User Management
  - Configuration
- **Process Execution**
  - Remote execution of processes
  - Scheduling processes
  - Exporting processes as Web services
- **Reports**
  - Using RapidMiner Processes as Web services in reports